

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 5, 2005, 00:24:30 ; Search time 4384.22 Seconds  
(without alignments)  
983.645 Million cell updates/sec

Title: US-10-691-590-1

Perfect score: 458

Sequence: 1 RREQEESSEETFGFQQVK.....FLNHKQNTNVKFTVKASAY 89

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command-line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US10691590/runat\_02112005\_120021\_13145/app\_query.fasta\_1.462  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10691590\_@CGN\_1\_1\_3851\_@runat\_02112005\_120021\_13145 -NCPU=3  
-NO MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_hgt.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	458	100.0	3318	8	TCCSV
2	339	74.0	1867	6	A20606 67kD protei
3	190	41.5	1248	8	AJ551425 Lens culi
4	185	40.4	1248	8	AJ626897 Pisum sat

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

5	185	40.4	1248	8	AJ626898
6	185	40.4	1596	8	PSVICK
7	183	40.0	907	8	PEAVIC2
8	182.5	39.8	1732	8	AY102931
9	181	39.5	114452	8	AC148289
10	177.5	38.8	2847	8	AB019195
11	177	38.6	1257	8	LCU551424
12	175	38.2	1690	8	LCVICNC
13	175	38.2	2777	8	COTSPA
14	175	38.2	7876	12	SYNAGABA
15	174.5	38.1	2057	8	AF066055
16	174	38.0	1743	8	LEVICN
17	172.5	37.7	1693	8	LCOVICN
18	170	37.1	1741	8	LCVICNE
19	169	36.9	1561	8	VFVIC
20	169	36.9	4239	8	PSVICIL
21	168.5	36.8	5327	8	VFVICG
22	168	36.7	1742	8	LCVICNA
23	166	36.2	1433	8	PSVIC
24	165	36.0	1690	8	LVNIC
25	164	35.8	1548	8	VNVICLN
26	164	35.8	1965	8	COTSPC
27	161	35.2	3736	8	PSCVCA
28	159	34.7	1858	8	VNCONVN
29	157.5	34.4	314	8	LES270964
30	156.5	34.2	1418	8	AY581851
31	156.5	34.2	1930	6	BD172109
32	156.5	34.2	1930	6	AR257463
33	156.5	34.2	1930	6	BD107898
34	156.5	34.2	2032	6	AR257469
35	156.5	34.2	2032	6	AX148742
36	156.5	34.2	2032	6	AX155331
37	156.5	34.2	2032	6	ARQARAHI
38	156.5	34.2	2041	6	AR257470
39	156.5	34.2	2674	8	AF432231
40	156	34.1	2069	8	PSA276875
41	156	34.1	2212	8	LCU276683
42	154.5	33.7	1978	6	AX155332
43	153.5	33.7	1035	8	AY581850
44	152.5	33.3	1595	8	OSU45322
45	151.5	33.1	3636	8	SOYBSP

#### ALIGNMENTS

##### RESULT 1

TCCSV	TCCSV	3318 bp	DNA	linear	PLN 03-DEC-1993
LOCUS	T.cacao csv gene for seed vicilin.				
DEFINITION	X62625 S38078				
ACCESSION	X62625.1 GI:21910				
VERSION	X62625.1				
KEYWORDS	csv gene; seed protein; vicilin.				
SOURCE	Theobroma cacao (cacao)				
ORGANISM	Theobroma cacao				
REFERENCE	1 (bases 1 to 3318)				
AUTHORS	McHenry, L. and Fritz, P.J.				
TITLE	Comparison of the structure and nucleotide sequences of vicilin				
JOURNAL	genes of cocoa and cotton raise questions about vicilin evolution				
MEDLINE	Plant Mol. Biol. 18 (6), 1173-1176 (1992)				
PUBMED	92288309				
REFERENCE	2 (bases 1 to 3318)				
AUTHORS	McHenry, L.				
TITLE	Direct Submission				
JOURNAL	Submitted (15-OCT-1991) L. McHenry, Pennsylvania State University,				
COMMENT	111 Borland Lab, University Park, PA 16802, USA				
FEATURES	See also X62626 Overlap of sequenced fragments. Location/Qualifiers				

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;  
Theobroma.

1 (bases 1 to 3318)

McHenry, L. and Fritz, P.J.

Comparison of the structure and nucleotide sequences of vicilin

genes of cocoa and cotton raise questions about vicilin evolution

Plant Mol. Biol. 18 (6), 1173-1176 (1992)

92288309

1600151

2 (bases 1 to 3318)

McHenry, L.

Direct Submission

Submitted (15-OCT-1991) L. McHenry, Pennsylvania State University,

111 Borland Lab, University Park, PA 16802, USA

See also X62626

Overlap of sequenced fragments.

Location/Qualifiers

source	1. 3318		US-10-691-590-1 (1-89) x TCCSV (1-3318)	
	/organism="Theobroma cacao"			
TATA_signal	/mol_type="genomic DNA"		QY 1 ArgArgGluGlnGluGluSerGluGluThrPheGlyGluPheGlnGlnVallys 20	
	/db_xref="taxon:3641"			
gene	/clone="pHD5P1.7, pHD5P1.2, pHD5E4.5"		Db 2312 AGAAGAGAACAAAGAGAGTCAGAGAGAGACATTTGGAGAAATCCAGCAGGTCAA 2371	
	/tissue_type="leaves"			
CDS	/clone_lib="cocoa genomic library in lambda GEM II"		QY 21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40	
	/dev_stage="mature tree"			
sig_peptide	543. .549		Db 2372 GCCCCAATTGTACCTGGTGACGCTTTTGTAGCCCGGAGGCGCATGCAGTTACATTCATT 2431	
	/note="csv gene"			
mat_peptide	610. .2581		QY 41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnGln 60	
	/gene="csv"			
exon	join(610. .1240,1331. .1506,1597. .1677,1780. .2064,		Db 2432 GCATCCAAGACCAGCCCTGATGCACTTGGCTTGGACTCAACGCCAGAACACACAG 2491	
	2177. .2581)			
intron	/gene="csv"		QY 61 ArgIlePheLeuAlaGlyArgProPhePheLeuAsnHisLysGlnAsnThrAsnValIle 80	
	/codon_start=1			
exon	/product="vicilin"		Db 2492 AGAATTTTCTTGCAGGTAGGCCTTTTCTGATCAACAAGCAAAACACAAATGTCATC 2551	
	/protein_id="CAA44493.1"			
intron	/db_xref="GI:21911"		QY 81 LysPheThrValLysAlaSerAlaTyr 89	
	/db_xref="GOA:Q43358"			
exon	/translation="MVISKSPFIVLIFSLLSFALLSCGVSAYGRKQYERDPQQYEQ		Db 2552 AAATTCAGTGAAGACCTCCGCTTAT 2578	
	CORCESEATEBEREQCEQRCEREYKEQREELQRYQQCQRCQEQEQQQQREQ			
intron	QCFKILORFAENSPLKGINDYRLAMFEANPTFLPHCDAAEAIYFTNKGKGTTFV		RESULT 2	
	THEKESINVGRTVVSVPAGSTVTVVSQDQEKLTIAVLALPVNSPGKYLEFPFAGN		A20606	
exon	NKPSYYGAFSYEVLETVFNTQREKLEELBEQRCQKQEQGQGMFRKAKPEQIRAI		LOCUS 67kD protein. 1867 bp DNA linear PAT 27-JUL-1994	
	QOATSPRRHGERLAINLLSQSPVYSNQNGRFFEACPEDFSQFQNMMDVAVSAFKLNQ			
intron	AI FVPHYNSKATFVVVTDGYGQAQMACPHLSRQSQSGRQDRREQEESSEETFG		DEFINITION A20606	
	EFOQVKAPLSPGDVFVAPAGHAVTFEASQDQPLNAVAFGLNAQNQRIFLAGRPFLN			
exon	HKQNTNVIKFTVKASAY"		ACCESSION A20606.1 GI:579221	
	610. .681			
intron	/gene="csv"		VERSION A20606.1	
	join(682. .1240,1331. .1506,1597. .1677,1780. .2064,			
exon	2177. .2578)		KEYWORDS	
	/gene="csv"			
intron	/product="vicilin"		SOURCE	
	4610. .1240			
exon	/gene="csv"		ORGANISM	
	/number=1			
intron	1241. .1330		Theobroma cacao (cacao)	
	/gene="csv"			
exon	/number=1		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
	1331. .1506			
intron	/gene="csv"		Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;	
	/number=2			
exon	1507. .1596		rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;	
	/gene="csv"			
intron	/number=2		Theobroma.	
	1597. .1677			
exon	/gene="csv"		REFERENCE	
	/number=3			
intron	1678. .1779		1 (bases 1 to 1867)	
	/gene="csv"			
exon	/number=3		AUTHORS	
	1780. .2064			
intron	/gene="csv"		TITLE	
	/number=4			
exon	2065. .2176		JOURNAL	
	/gene="csv"			
intron	/number=4		FEATURES	
	2177. .25282			
exon	/number=5		source	
	2856. .2861			
polyA_signal	/gene="67kD protein"		gene	
	/codon_start=1			
ORIGIN	/protein_id="CAA01512.1"		CDS	
	/db_xref="GI:579222"			
Alignment Scores:	/translation="MVISKSPFIVLIFSLLSFALLSCGVSAYGRKQYERDPQQYEQ		Pred. No.:	
	CORCESEATEBEREQCEQRCEREYKEQREELQRYQQCQRCQEQEQQQQREQ			
Pred. No.:	QCFKILORFAENSPLKGINDYRLAMFEANPTFLPHCDAAEAIYFTNKGKGTTFV		7.33e-32	
	THEKESINVGRTVVSVPAGSTVTVVSQDQEKLTIAVLALPVNSPGKYLEFPFAGN			
Score:	NKPSYYGAFSYEVLETVFNTQREKLEELBEQRCQKQEQGQGMFRKAKPEQIRAI		1867	
	QOATSPRRHGERLAINLLSQSPVYSNQNGRFFEACPEDFSQFQNMMDVAVSAFKLNQ			
Percent Similarity:	AI FVPHYNSKATFVVVTDGYGQAQMACPHLSRQSQSGRQDRREQEESSEETFG		339.00	
	EFOQVKAPLSPGDVFVAPAGHAVTFEASQDQPLNAVAFGLNAQNQRIFLAGKKNLVR			
Best Local Similarity:	QMDSEAKELSPGVPSKLVNDIFNPNDESIFMSFSQSRQRDRDRGNPLASILDARLEF		Matches: 66	
	"			
Query Match:	100.00%		Conservative: 1	
	100.00%			
DB:	8		Mismatch: 0	
	0			
DB:	0		Indels: 0	
	0			
DB:	0		Gaps: 0	
	0			

US-10-691-590-1 (1-89) x A20606 (1-1867)

QY 1 ArgArgGluGlnGluGluSerGluGluThrPheGlyGluPheGlnValValys 20  
 Db 1322 AGAAGAGAACAAAGAGAGAGTTCAGAGAGGACATTTTGAGAAATTCACAGGTCAAA 1381  
 QY 21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40  
 Db 1382 GCCCATTTGTCACCTGGTGACGCTTTGTAGCCCCGGCAGGCCATCAGTTACATCTTT 1441  
 QY 41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnGln 60  
 Db 1442 GCATCAAGACACAGCCCTGAATGCAGTTGCGTTTGCACTCAACGCCCAAGCAACCCAG 1501  
 QY 61 ArgIlePheLeuAlaGlyArg 67  
 Db 1502 AGAATTTTCTTCAGGGAAA 1522

# RESULT 3

LCU551425 1248 bp mRNA linear PLN 19-DEC-2003  
 LOCUS  
 DEFINITION Lens culinaris vicilin partial mRNA for allergen Len c 1.0102.  
 ACCESSION AJ551425  
 VERSION AJ551425.1 GI:29539110  
 KEYWORDS allergen Len c 1.0102.  
 SOURCE Lens culinaris (lentil)  
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Lens.

REFERENCE 1 Lopez-Torrejon, G., Salcedo, G., Martin-Esteban, M., Diaz-Perales, A., Pascual, C.Y. and Sanchez-Monge, R.  
 Len c 1, a major allergen and vicilin from lentil seeds: Protein isolation and cDNA cloning  
 J. Allergy Clin. Immunol. 112 (6), 1208-1215 (2003)

PUBMED 14657885

REFERENCE 2 (bases 1 to 1248)

Salcedo, G.

Direct Submission

TITLE Submitted (27-MAR-2003) Salcedo G., Biotechnology, E.T.S.

JOURNAL Agronomos, Ciudad Universitaria. Madrid, 28040, SPAIN

## FEATURES

source

1..1248  
 /organism="Lens culinaris"  
 /mol\_type="mRNA"  
 /cultivar="Guardena"  
 /db\_xref="taxon:3864"  
 /country="Spain"

<1..1248

/function="seed storage protein"

/codon\_start=1

/evidence=experimental

/product="allergen Len c 1.0102"

/protein\_id="CA887731.1"

/db\_xref="GI:29539111"

/db\_xref="GOA:Q84UI0"

/db\_xref="UniProt/TREMBL:Q84UI0"

/translations="SRSDQNPPIFKSNRFTIYENENGHIRLLQKFDKRSKIFENLQ  
 NYRLLEYSKPHTLFLPQYTDADFIIVLSGKATLVLSNDNRNSFLREGDTIKLPA  
 GTTAYLANRDNDLRLVLDIAI PVNPGQLQSFLLSGTQNPQFLSGFNKSLLEAAFN  
 TDYEETKVLLEQEOEPHRRSLRDRRQIBNKENVIVKVSREQIKLSNKAASSKK  
 SVSESGPNLRSRNPISYNSKFGKFEITPEKNQQLDLDIFVNSVIEKGSLLLPNY  
 NSRAIVIVITVEGKGYPQLVQNRNENDEEBEETSTQVQRYAKLSFGDVF  
 VPAIGHVPAVNASDLNLIGFGINAKNQRNFLAGEDNVI SQIQRPVKELAPFGSSR  
 EVDRLLTNQKQSFANAPLQIE"

mat\_peptide

1..1245

/product="allergen Len c 1.0102"

/evidence=experimental

## ORIGIN

Alignment Scores:

Pred. No.: 1.12e-13 Length: 1248  
 Score: 190.00 Matches: 42  
 Percent Similarity: 64.10% Conservative: 8  
 Best Local Similarity: 53.85% Mismatches: 18  
 Query Match: 41.48% Indels: 10  
 DB: Gaps: 2

US-10-691-590-1 (1-89) x LCU551425 (1-1248)

QY 3 GluGlnGluGluSerGluGluThrPheGlyGluPheGlnValValysAlaPro 22  
 Db 922 GACGATGAGAGAGAACAAAGAGAGACACACAGTGCAGAGGTATAGGGCTAAG 981  
 QY 23 LeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePheAlaSer 42  
 Db 982 TTGTCTCCAGGTGATGTTTTTGTGGTTCACGAGGTCCAGTTCGCAATAATGCTCC 1041  
 QY 43 LysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnGlnArgIle 62  
 Db 1042 TCAGAT-----CTCAATCTAATGGATTTGGTATCAATGCCAAAACAATCAGAGAAAC 1095  
 QY 63 PheLeuAlaGlyArgProphePheLeuAsnHisGlnAsnThrAsnValIle 80  
 Db 1096 TTTCTGCGCAGGC-----GAGGAGGACAATGTGATA 1125

# RESULT 4

AJ626897 1248 bp mRNA linear PLN 05-FEB-2004  
 LOCUS  
 DEFINITION Pisum sativum partial mRNA for Vicilin, allergen pis s 1.0101.

ACCESSION AJ626897

VERSION AJ626897.1 GI:42414626

KEYWORDS allergen pis s 1.0101; Vicilin.

SOURCE Pisum sativum (pea)

ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.

## REFERENCE

AUTHORS Sanchez-Monge, R., Lopez-Torrejon, G., Pascual, C.Y., Varela, J., Martin-Esteban, M. and Salcedo, G.

TITLE Vicilin and convicilin are potential major allergens from pea seeds

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1248)

Salcedo, G.

Direct Submission

TITLE Submitted (04-FEB-2004) Salcedo G., Biotechnology microbiana.

JOURNAL Bioquimica, E.T.S. Ingenieros Agronomos, Ciudad Universitaria,

28040 Madrid, SPAIN

## FEATURES

source

1..1248  
 /organism="Pisum sativum"  
 /mol\_type="mRNA"  
 /cultivar="Rommel"  
 /db\_xref="taxon:3888"

<1..1248

/function="seed storage protein"

/note="allergen pis s 1.0101"

/codon\_start=1

/evidence=experimental

/product="Vicilin"

/protein\_id="CAF25232.1"

/db\_xref="GI:42414627"

/translations="SRSDQNPPIFKSNRFTIYENENGHIRLLQKFDKRSKIFENLQ  
 NYRLLEYSKPHTLFLPQYTDADFIIVLSGKATLVLSNDNRNSFLREGDTIKLPA  
 GTTAYLANRDNDLRLVLDIAI PVNPGQLQSFLLSGTQNPQFLSGFNKSLLEAAFN  
 TDYEETKVLLEQEOEPHRRSLRDRRQIBNKENVIVKVSREQIKLSNKAASSKK  
 SVSESGPNLRSRNPISYNSKFGKFEITPEKNQQLDLDIFVNSVIEKGSLLLPNY  
 NSRAIVIVITVEGKGYPQLVQNRNENDEEBEETSTQVQRYAKLSFGDVF  
 VPAIGHVPAVNASDLNLIGFGINAKNQRNFLAGEDNVI SQIQRPVKELAPFGSSR  
 EVDRLLTNQKQSFANAPLQIE"

1..1245

/product="allergen pis s 1.0101"

/evidence=experimental

ORIGIN

```

Alignment Scores:
Pred. No.: 4,64e-13 Length: 1248
Score: 185.00 Matches: 41
Percent Similarity: 64.10% Conservative: 9
Best Local Similarity: 52.56% Mismatches: 18
Query Match: 40.39% Indels: 10
DB: 8 Gaps: 2

US-10-691-590-1 (1-89) x AJ626897 (1-1248)
Qy 3 GluGlnGluGluSerGluGluThrPheGlyGluPheGlnGlnVallysalaPro 22
Db 922 GACACGAGAGAGCAACAGAGAGAGCAACAAAGTCAACTGTATAGAGCTAAG 981
Qy 23 LeuSerProGlyAspValPheValalaProAlaGlyHisAlaValThrPhePheAlaSer 42
Db 982 TTGCTCCAGGTGATGTTTGTGATTCAGCAGGTCAACCCGTTGCCATAAATGCCTCC 1041
Qy 43 LysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnGlnArgile 62
Db 1042 TCAGAT-----CTCAATCTGATGTTGTATCAATGCCGAGAACACGAGAGAAC 1095
Qy 63 PheLeuAlaGlyArgProPhePheLeuAsnHisLysGlnAsnThrAsnValille 80
Db 1096 TTCCTTGACGT-----GAGGAAGACAAATGTCATA 1125

RESULT 5
AJ626898
LOCUS Pisum sativum partial mRNA for Vicilin, allergen pis s 1.0102.
DEFINITION AJ626898
ACCESSION AJ626898
VERSION AJ626898.1 GI:42414628
KEYWORDS allergen pis s 1.0102; Vicilin.
SOURCE Pisum sativum (pea)
ORGANISM Pisum
REFERENCE 1
AUTHORS Sanchez-Monge,R., Lopez-Torrejón,G., Pascual,C.Y., Varela,J.,
TITLE Martin-Esteban,M. and Salcedo,G.
JOURNAL Vicilin and convicilin are potential major allergens from pea seeds
REFERENCE 2 (bases 1 to 1248)
AUTHORS Unpublished
TITLE Salcedo,G.
JOURNAL Direct Submission
SUBMITTED (04-FEB-2004) Salcedo G., Biotechnology microbiana.
BIOQUIMICA E.T.S. Ingenieros Agronomos, Ciudad Universitaria,
28040 Madrid, SPAIN
FEATURES
Location/Qualifiers
1..1248
/organism="Pisum sativum"
/mol_type="mRNA"
/cultivar="Rommel"
/db_xref="taxon:3888"
<1..1248
/function="seed storage protein"
/notes="allergen pis s 1.0102"
/codon_start=1
/evidence=experimental
/product="Vicilin"
/protein_id="CAF25233.1"
/db_xref="GI:42414629"
</db_xref>
CDS
translation="SRSDQENPFIFKSNRFQTLYENENGHIRLLQKRRSKIFENLQ
NYRLVYKSPRTFLPQCTDADFLVLSGKATLVLSKNDNRNSFLNLRGDTIKLPA
GTIAYLANRDDNEDLVLTITPVNKPQQLQSLGCTONQPSILSGFSKNILEAFN
TNYEEKVLKQOEPOPHRRSLKDRRIENEVIVKVSREQHELSKNKAKSSKK
SVSSESGPFLNRSNPYKNKGKFEITPEKNQQLQDLIDFVNSVDIKESGLLUPNY
NSRAIVIVTEKGDFELVGNQENKDEKEEETSKQVLYRAKLSPGDVF
VIPAGHPVAINASSDLNLIGFGINAENNRNFLAGEDNVISQVERPVKELAPGGSH
EVDRLLNKQKQSYFANAQPLQRE"
ORIGIN

```

```

Alignment Scores:
Pred. No.: 4,64e-13 Length: 1248
Score: 185.00 Matches: 41
Percent Similarity: 64.10% Conservative: 9
Best Local Similarity: 52.56% Mismatches: 18
Query Match: 40.39% Indels: 10
DB: 8 Gaps: 2

US-10-691-590-1 (1-89) x AJ626898 (1-1248)
Qy 3 GluGlnGluGluSerGluGluThrPheGlyGluPheGlnGlnVallysalaPro 22
Db 922 GACACGAGAGAGCAACAGAGAGAGCAACAAAGTCAACTGTATAGAGCTAAG 981
Qy 23 LeuSerProGlyAspValPheValalaProAlaGlyHisAlaValThrPhePheAlaSer 42
Db 982 TTGCTCCAGGTGATGTTTGTGATTCAGCAGGTCAACCCGTTGCCATAAATGCCTCC 1041
Qy 43 LysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnGlnArgile 62
Db 1042 TCAGAT-----CTCAATCTGATGTTGTATCAATGCCGAGAACACGAGAGAAC 1095
Qy 63 PheLeuAlaGlyArgProPhePheLeuAsnHisLysGlnAsnThrAsnValille 80
Db 1096 TTCCTTGACGT-----GAGGAAGACAAATGTCATA 1125

RESULT 6
PSVICK
LOCUS Pisum sativum vick gene.
DEFINITION X67429
ACCESSION X67429.1 GI:297169
VERSION X67429.1
KEYWORDS 47 kDa protein; seed storage protein; vicilin; vick gene.
SOURCE Pisum sativum (pea)
ORGANISM Pisum
REFERENCE 1 (bases 1 to 1596)
AUTHORS Thompson,A.J., Bown,D.P., Yaish,S. and Gatehouse,J.A.
TITLE Differential expression of seed storage protein genes in the pea.
JOURNAL Biochem. Physiol. Pflanz. 187, 1-12 (1991)
REFERENCE 2 (bases 1 to 1596)
AUTHORS Bown,D.P.
TITLE Direct Submission
SUBMITTED (02-JUN-1992) D.P. Bown, University of Durham, Dept. of
Biological Sciences, Science Site, South Rd, Durham, DH1 3LE, UK
FEATURES
Location/Qualifiers
1..1596
/organism="Pisum sativum"
/mol_type="mRNA"
/cultivar="Feltham first"
/db_xref="taxon:3888"
/clone="PLGI.63"
/tissue_type="cotyledon"
/clone_lib="plasmid pUC8"
/dev_stage="13-17 days post flowering"
1..1596
/gene="vick"
13..1329
/gene="vick"
/codon_start=1
/product="vicilin 47kd protein"
/protein_id="CAA47814.1"
/db_xref="GI:297170"
/db_xref="GOA:Q43626"
/db_xref="UniProt/TREMBL:Q43626"
</db_xref>
translation="MAATPIKPLMLLAFLASVCVSSRSDQENPFIFKSNRFQTLYE
gene
CDS

```















GenCore version 5.1.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 5, 2005, 00:20:21 ; Search time 568.826 Seconds  
(without alignments)  
926.218 Million cell updates/sec

Title: US-10-691-590-1

Perfect score: 458

Sequence: 1 RREQESESEETGFEQVK.....FLNHKQNTNWKFTVKASAY 89

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US10691590/runat\_02112005\_120021\_13134/app.query.fasta\_1.462  
-DB=N Geneseq 16Dec04 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10691590 @CGN\_1\_156 @runat\_02112005\_120021\_13134 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGIOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 16Dec04.\*

1: Geneseqn1980s.\*  
2: Geneseqn1990s.\*  
3: Geneseqn2000s.\*  
4: Geneseqn2001as.\*  
5: Geneseqn2001bs.\*  
6: Geneseqn2002as.\*  
7: Geneseqn2002bs.\*  
8: Geneseqn2003as.\*  
9: Geneseqn2003bs.\*  
10: Geneseqn2003cs.\*  
11: Geneseqn2003ds.\*  
12: Geneseqn2004as.\*  
13: Geneseqn2004bs.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	339	74.0	1867	2 AAQ20377	Sequence
2	159	34.7	562	13 ACN45981	Acn45981 Cotton pr
3	156.5	34.2	1930	2 AAZ06382	Aaz06382 Peanut al
4	156.5	34.2	2032	2 AAT76613	Aat76613 Peanut al
5	156.5	34.2	2032	4 AAF90339	Aaf90339 Peanut al

6	156.5	34.2	2032	8 ABX70603	Abx70603 Peanut Ar
7	156.5	34.2	2032	10 ADG27462	Adg27462 Peanut Ar
8	156.5	34.2	2041	4 AAS08537	Aas08537 DNA encod
9	154.5	33.7	498	13 ACN48047	Acn48047 Cotton pr
10	151	33.0	1949	2 AAT76612	Aat76612 Peanut al
11	151	33.0	1949	8 ABX70604	Abx70604 Peanut Ar
12	151	33.0	1949	10 ADG27463	Adg27463 Peanut Ar
13	149.5	32.6	1722	10 ADL18492	Adl18492 Maize glo
14	149.5	32.6	2003	10 ADL18494	Adl18494 Maize glo
15	143	31.2	1680	6 ABS55197	Abs55197 Glycine m
16	142	31.0	1920	2 AAV17563	Aav17563 Coding se
17	140.5	30.7	2171	2 AAV42311	Aav42311 Macadamia
18	140	30.6	1251	6 ABS55196	Abs55196 Glycine m
19	140	30.6	1254	9 ACC49562	Acc49562 Mature be
20	140	30.6	1278	9 ACC49561	Acc49561 FLAG-tag
21	140	30.6	1320	2 AAV17564	Aav17564 Coding se
22	140	30.6	1320	9 ACC49553	Acc49553 Glycine m
23	137	29.9	1632	6 ABS55198	Abs55198 Glycine m
24	137	29.9	1818	2 AAV17562	Aav17562 Coding se
25	134.5	29.4	1952	4 AAS08539	Aas08539 Anaphylac
26	134	29.3	2140	2 AAV42316	Aav42316 Macadamia
27	133.5	29.1	1350	12 ADJ44543	Adj44543 Plant cdn
28	131.5	28.7	2171	2 AAV42310	Aav42310 Macadamia
29	114	24.9	1581	10 ABX11299	Abx11299 Loblolly
30	114	24.9	1596	10 ABX16585	Abx16585 Loblolly
31	109	23.8	591	13 ACN50453	Acn50453 Cotton ma
32	109	23.8	659	13 ACN50255	Acn50255 Cotton no
33	109	23.8	2332	13 ADR63468	Adr63468 Cotton cd
34	108	23.6	564	13 ACN49247	Acn49247 Cotton pr
35	102	22.3	705	12 ADJ40042	Adj40042 Plant cdn
36	100	21.8	1924	2 AAV72243	Aav72243 G. max SB
37	99.5	21.7	1580	12 ADE80950	Ade80950 Cashew nu
38	99.5	21.7	1730	12 ADE80949	Ade80949 Cashew nu
39	87	19.0	559	13 ACN48185	Acn48185 Cotton pr
40	83.5	18.2	1600	4 AAD17529	Aad17529 Soybean G
41	78	17.0	1500	10 ADH89318	Adh89318 A. thalia
42	78	17.0	1500	12 ADG44114	Adg44114 A. thalia
43	77.5	16.9	1855	4 AAF90338	Aaf90338 Peanut al
44	76.5	16.7	338702	11 ACN44042	Acn44042 Human gen
45	75.5	16.5	669	10 ADD16431	Add16431 DNA (seqi

#### ALIGNMENTS

RESULT 1  
AAQ20377  
ID AAQ20377 standard; cdna; 1867 BP.  
XX  
XX AAQ20377;  
XX AC  
XX 16-APR-1992 (first entry)  
XX  
XX Sequence of 67 kd T. cacao protein cdna.  
XX  
XX Cocoa; flavour; vicilin; seed storage protein; ss.  
XX Theobroma cacao.  
XX  
XX Key Location/Qualifiers  
XX CDS 14..1714  
XX FT /\*tag= a  
XX  
XX WO9119801-A.  
XX  
XX 26-DEC-1991.  
XX  
XX 11-JUN-1990; 90GB-00013016.  
XX  
XX 11-JUN-1990; 90GB-00013016.  
XX  
XX (MRSC ) MARS UK LTD.  
XX  
XX PI Spencer ME, Hodge R, Deakin EA, Ashton S;













XX 14-DEC-2000; 2000US-0255619P.  
XX (DEIK/) DEIKMAN J.  
XX (FENG/) FENG P C C.  
XX (FINC/) FINCHER K L.  
XX (ZIEG/) ZIEGLER T E.  
XX Deikman J, Feng PCC, Fincher KL, Ziegler TE;  
XX WPI; 2004-479808/45.  
XX  
XX New isolated nucleic acid molecule that encodes a plant protein or its  
XX fragment, useful for isolating a variety of agronomically significant  
XX genes associated with plant growth, quality or yield, and as molecular  
XX tags to map genes.  
XX  
XX Claim 1; SEQ ID NO 2828; 34pp; English.  
XX  
XX The invention relates to 17880 cotton expressed sequence tags (ESTs;  
XX ACN45220-ACN63099). The ESTs were isolated from cDNA libraries generated  
XX from primed or non-primed seeds from variety DP50B, mature seeds from  
XX variety Coker 312 Boswell 96 Field, and androecium tissue, gynoeceum  
XX tissue, developing fibres, carpal walls and septa from variety  
XX Nucton33B. The invention also relates to substantially purified  
XX proteins or their fragments encoded by nucleic acid molecules of the  
XX invention, and to transformed plants having a nucleic acid construct  
XX comprising a nucleic acid of the invention. The cotton ESTs are useful as  
XX molecular tags to isolate genetic regions, to isolate genes, to map  
XX genes, to determine gene function and to determine whether genes are  
XX members of a particular gene family. The nucleic acid molecules may be  
XX used for isolating a variety of agronomically significant genes  
XX associated with plant growth, quality, yield, and could also serve as  
XX links in metabolic and catabolic pathways. The nucleic acid molecules are  
XX also useful for identifying genes important in initiating and maintaining  
XX seed germination or that may be used to mitigate stresses encountered  
XX during seed germination. The ESTs additionally enable the acquisition of  
XX promoters and cis-regulatory elements which will be useful to express  
XX agronomically significant genes in these tissues and/or other tissues,  
XX and also permits the acquisition of molecular markers useful in breeding  
XX schemes, genetic and molecular mapping, and in cloning of agronomically  
XX significant genes. The nucleic acid molecules are further useful for  
XX detecting the presence or level or pattern of a protein or mRNA and for  
XX detecting the presence or quantity of a protein by tissue printing. The  
XX present sequence represents a specifically claimed EST isolated from a  
XX cotton variety DP50B primed seed cDNA library (Lib3825). The sequence  
XX data for this patent did not form part of the printed specification, but  
XX was obtained in electronic format directly from the US patent office at  
XX seqdata.uspto.gov/sequence.html?DocID=US20040123340  
XX  
XX Sequence 498 BP; 156 A; 125 C; 74 G; 143 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 1,01e-12 Length: 498  
Score: 154.50 Matches: 34  
Percent Similarity: 66.20% Conservatives: 13  
Best Local Similarity: 47.89% Mismatches: 15  
Query Match: 33.73% Indels: 9  
DB: 13 Gaps: 2

US-10-691-590-1 (1-89) x ACN48047 (1-498)

Qy 17 GlnGlnValLysAlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAla 36  
Db 498 AAAAAAGTGAGAGCGCGTTATCACCGGTGACATATTGTAGTCCAGCAGCCATCCA 439  
Qy 37 ValThrPhePheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeu---Asn 55  
Db 438 GTCACTATTGTTGCTCCCAAAACAAACCTAGATTGTTCGGTTTGACATTACAC 379  
Qy 56 AlaGlnAsnAsnGlnArgIlePheLeuAlaGlyArgPropPhePheLeuAsnHisLysGln 75  
Db 378 GGCCAAACACAAAGAGATTTTGTGGTTGGGAA----- 343

Qy 76 AsnThrAsnValIleLysPheThrValLysAla 86  
Db 342 ACCACCAATGTTAGACAATGGTTCGCCAGGCC 310  
RESULT 10  
AAT76612  
ID AAT76612 standard; cDNA to mRNA; 1949 BP.  
XX AAT76612;  
XX  
XX 17-OCT-2003 (revised)  
XX 29-DEC-1997 (first entry)  
XX Peanut allergen Ara hi cDNA clone p17.  
XX  
XX Peanut; seed storage protein; allergen; allergy; hypersensitivity;  
XX vaccine; anaphylactic shock; immunotherapy; therapy; monoclonal antibody;  
XX ELISA; analysis; Ara hi; ds.  
XX  
XX Arachis hypogaea; strain Florunner.  
XX  
XX Key Location/Qualifiers  
XX CDS 3..1847  
XX sig\_peptide /\*tag= a  
XX FT 3..68  
XX mat\_peptide /\*tag= b  
XX FT 69..1844  
XX polyA\_signal /\*tag= c  
XX FT 1918..1923  
XX /\*tag= d  
XX  
XX WO9724139-A1.  
XX  
XX 10-JUL-1997.  
XX  
XX 23-SEP-1996; 96WO-US015222.  
XX  
XX 29-DEC-1995; 95US-0009455P.  
XX 04-MAR-1996; 96US-00610424.  
XX  
XX (UYAR-) UNIV ARKANSAS.  
XX  
XX Burks AW, Helm RM, Cockrell G, Stanley JS, Bannon GA;  
XX WPI; 1997-363453/33.  
XX P-PSDB; AAW22149.  
XX  
XX Peanut allergens Ara hi and Ara hII - used for vaccination and in two-  
XX site monoclonal antibody based ELISA.  
XX  
XX Claim 31; Page 183-185; 354pp; English.  
XX  
XX This cDNA clone, designated p17, codes for the major peanut allergen Ara  
XX hi (AAW22149), which has multiple IgE binding epitopes (see AAW24165-87).  
XX It was amplified from peanut seed cDNA using a primer (see AAT76616)  
XX based on an isolated Ara hi peptide (see AAW24206). The sequence shows  
XX significant homology with the vicilin family of seed storage proteins of  
XX other legumes. The gene is capable of producing a protein product in  
XX prokaryotic cells that is recognised by serum IgE from a large proportion  
XX of individuals with peanut hypersensitivity. Ara hi and Ara hII (see  
XX AAW24164) can be used to raise monoclonal antibodies which are used in a  
XX specific two-site MAb ELISA for the detection of Ara hi or Ara hII  
XX (claimed). IgE-binding Ara antigen epitopes may be used in vaccines to  
XX protect against allergic reactions to peanut allergens, e.g. anaphylactic  
XX shock. (Updated on 17-OCT-2003 to standardise OS field)  
XX  
XX Sequence 1949 BP; 599 A; 455 C; 517 G; 378 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 2,21e-11 Length: 1949  
Score: 151.00 Matches: 36  
Percent Similarity: 63.89% Conservatives: 10

Best Local Similarity: 50.00% Mismatches: 18  
Query Match: 32.97% Indels: 8  
DB: 2 Gaps: 3

US-10-691-590-1 (1-89) x AAT76612 (1-1949)

QY 1 ArgATGGluGln-----GluGluGluSerGluGluGluThrPheGlyGluPheGln 17  
Db 1413 CGCGCGGAACAAGAGTGGGAAGAAGAGGAGGAAGTGAAGAGAGGAGGAAGTAACAGA 1472  
QY 18 GlnValLys-----AlaProLeuSerProGlyAspValPheValAlaProAlaGly 34  
Db 1473 GAGGTGCGTAGGTACACAGCGAGGTTGAAGGAAGCGATGTGTTTCATCATGCCAGCAGCT 1532  
QY 35 HisAlaValThrPhePheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeu 54  
Db 1533 CATCCAGTAGCCATCAACGCTTCTCCGAA-----CTCCATCTGCTTGGCTTCGGTATC 1586  
QY 55 AsnAlaGlnAsnAsnGlnArgIlePheLeuAlaGly 66  
Db 1587 AACGCTGAAACAAACACACAGAAATCTTCTTGCAGGT 1622

## RESULT 11

ABX70604  
ID ABX70604 standard; cDNA; 1949 BP.

XX AC ABX70604;

XX DT 26-MAR-2003 (first entry)

XX DE Peanut Ara h1 cDNA clone P17.

XX KW Peanut; ss; allergy; Ara h1; Ara h2; Ara h3; gene; IgE binding site;  
XX KW anaphylactic food allergen; antiallergenic; vaccine; wound healing.

XX OS Arachis hypogaea.

XX PN WO200274250-A2.

XX PD 26-SEP-2002.

XX PF 18-MAR-2002; 2002WO-US0009108.

XX PR 16-MAR-2001; 2001US-0276822P.

XX PR 18-MAR-2002; 2002US-00276822.

XX PA (PANA-) PANACEA PHARM.

XX PI Caplan M, Sosin H, Sampson H, Bannon GA, Burks WA, Cockrell G;  
XX PI Compadre CM, Connaughton C, Helm RM, King NE, Kopper RA, Maleki SJ;  
XX PI Rabjohn PA, Shin DS, Stanley JS;

XX DR WPI; 2003-018765/01.

XX DR P-PSDB; ABUS2413.

XX PT New modified anaphylactic food allergen, useful for preventing or  
XX PT treating allergic reactions associated with e.g. anaphylactic allergens.

XX PS Example 3; Fig 11; 300pp; English.

XX CC The invention relates to a modified anaphylactic food allergen has an  
XX CC amino acid sequence that is substantially identical to that of natural  
XX CC anaphylactic food allergen, except for a cysteine residue that has been  
XX CC modified so that it cannot participate in the disulphide bond. The  
XX CC modification may also comprise mutation of the IgE binding sites to  
XX CC reduce allergenicity. Also included are: (1) a method of making a  
XX CC modified anaphylactic food allergen; (2) a nucleotide molecule encoding  
XX CC or for causing a site specific mutation in the modified anaphylactic food  
XX CC allergen; (3) a transgenic plant or animal expressing the modified  
XX CC anaphylactic food allergen; (4) a method of treating an individual by  
XX CC reducing the clinical response to a natural anaphylactic food allergen;  
XX CC and an isolated fragment of peanut allergen Ara h 1. The modified  
XX CC anaphylactic food allergen is useful for preventing or treating allergic

CC reactions associated with any natural allergen such as food, insect,  
CC rubber or preferably anaphylactic allergens. It is also useful for  
CC treating wounds in mammals such as bovine, canine, feline, caprine,  
CC ovine, porcine, murine or equine species. The present sequence is a cDNA  
CC encoding a peanut allergen (e.g. Ara h1, h2 or h3)

SQ Sequence 1949 BP; 599 A; 455 C; 517 G; 378 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 2,21e-11 Length: 1949  
Score: 151.00 Matches: 36  
Percent Similarity: 63.89% Conservative: 10  
Best Local Similarity: 50.00% Mismatches: 18  
Query Match: 32.97% Indels: 8  
DB: 8 Gaps: 3

US-10-691-590-1 (1-89) x ABX70604 (1-1949)

QY 1 ArgATGGluGln-----GluGluGluSerGluGluGluThrPheGlyGluPheGln 17  
Db 1413 CGCGCGGAACAAGAGTGGGAAGAAGAGGAGGAAGTGAAGAGAGGAGGAAGTAACAGA 1472

QY 18 GlnValLys-----AlaProLeuSerProGlyAspValPheValAlaProAlaGly 34  
Db 1473 GAGGTGCGTAGGTACACAGCGAGGTTGAAGGAAGCGATGTGTTTCATCATGCCAGCAGCT 1532

QY 35 HisAlaValThrPhePheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeu 54  
Db 1533 CATCCAGTAGCCATCAACGCTTCTCCGAA-----CTCCATCTGCTTGGCTTCGGTATC 1586

QY 55 AsnAlaGlnAsnAsnGlnArgIlePheLeuAlaGly 66  
Db 1587 AACGCTGAAACAAACACACAGAAATCTTCTTGCAGGT 1622

## RESULT 12

ADG27463  
ID ADG27463 standard; cDNA; 1949 BP.

XX AC ADG27463;

XX DT 26-FEB-2004 (first entry)

XX DE Peanut Ara h1 cDNA clone P17.

XX KW Peanut; plant; ss; gene; allergen; Ara h1; Ara h2; Ara h3;  
XX KW glycinin A2B1a; Jug n1; antiallergic; vulnerary;  
XX KW anaphylactic food allergen; IgE; allergy; wound.

XX OS Arachis hypogaea.

XX XX US2003202980-A1.

XX XX 30-OCT-2003.

XX XX 18-MAR-2002; 2002US-00100303.

XX XX 29-DEC-1995; 95US-0009455P.

XX XX 23-SEP-1996; 96US-00717933.

XX XX 31-JAN-1998; 98US-0073283P.

XX XX 13-FEB-1998; 98US-0074590P.

XX XX 13-FEB-1998; 98US-0074624P.

XX XX 13-FEB-1998; 98US-0074633P.

XX XX 27-JUN-1998; 98US-00106872.

XX XX 27-AUG-1998; 98US-00141220.

XX XX 13-NOV-1998; 98US-00191593.

XX XX 29-JAN-1999; 99US-00240557.

XX XX 29-JAN-1999; 99US-00241101.

XX XX 11-FEB-1999; 99US-00248673.

XX XX 11-FEB-1999; 99US-00248674.

XX XX 02-MAR-1999; 99US-0122450P.

XX XX 02-MAR-1999; 99US-0122452P.

XX XX 02-MAR-1999; 99US-0122560P.

XX XX 02-MAR-1999; 99US-0122565P.

```
PR 02-MAR-1999; 99US-0122566P.
PR 11-MAR-1999; 99US-00267719.
PR 28-JAN-2000; 2000US-00494096.
PR 16-MAR-2001; 2001US-0276822P.
XX
PA (CAPL/) CAPLAN M J.
PA (SOSI/) SOSIN H B.
PA (SAMP/) SAMPSON H.
PA (BANN/) BANNON G A.
PA (BURK/) BURKS A W.
PA (COCK/) COCKRELL G.
PA (COMP/) COMPADRE C M.
PA (CONN/) CONNAUGHTON C.
PA (HELM/) HELM R M.
PA (KING/) KING N E.
PA (KOPP/) KOPPER R A.
PA (MALE/) MALEKI S J.
PA (RABJ/) RABJOHN P A.
PA (SHIN/) SHIN D S.
PA (STAN/) STANLEY J S.
XX
XX Caplan MJ, Sosin HB, Sampson H, Bannon GA, Burks AW, Cockrell G;
PI Compadre CM, Connaughton C, Helm RM, King NE, Kopper RA, Maleki SJ;
PI Rabjohn PA, Shin DS, Stanley JS;
XX
XX WPI: 2003-875632/81.
DR P-PSDB; ADG27465.
XX
XX New modified anaphylactic food allergen comprising a cysteine residue
PT which has been modified so that it cannot participate in the disulfide
PT bond, useful for treating allergic reactions or wounds.
XX
XX Example 4; SEQ ID NO 6; 194pp; English.
XX
XX The invention relates to a modified anaphylactic food allergen whose
CC amino acid sequence is substantially identical to that of a natural
CC anaphylactic food allergen. The natural anaphylactic food allergen
CC includes at least one cysteine residue that participates in a disulphide
CC bond when the natural anaphylactic food allergen is in its native
CC conformation, except that the cysteine residue has been modified so that
CC it cannot participate in the disulphide bond. Also included are a method
CC of making a modified anaphylactic food allergen, a nucleotide molecule
CC encoding a modified anaphylactic food allergen defined above, a
CC nucleotide molecule for causing a site specific mutation in a gene
CC encoding a natural anaphylactic food allergen, a transgenic plant or
CC animal expressing a modified anaphylactic food allergen defined above, a
CC method of treating an individual by reducing the clinical response to a
CC natural anaphylactic food allergen by administering a modified
CC anaphylactic food allergen and an isolated fragment of peanut allergen
CC Ara h 1, comprising at least 10 consecutive amino acids of ADG27464 or
CC ADG27465. About 10-17% of the amino acids have been modified in at least
CC one IGE epitope or all the IGE epitopes recognised when the natural
CC anaphylactic food allergen is contacted with serum IGE from individual (s)
CC allergic to the natural anaphylactic food allergen. The invention
CC discloses Peanut allergens Ara h1, Ara h2, Ara h3 (and their encoding
CC cDNAs), Soybean Glycinin A2B1a and IGE-binding epitopes of the English
CC walnut allergen Jug ni. The modified anaphylactic food allergen can be
CC used for treating allergic reactions or wounds. The present sequence
CC encodes a Peanut allergen of the invention.
XX
SQ Sequence 1949 BP; 599 A; 456 C; 516 G; 378 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Pred. No.: 2.21e-11 Length: 1949
XX Score: 151.00 Matches: 36
XX Percent Similarity: 63.89% Conservative: 10
XX Best Local Similarity: 50.00% Mismatches: 18
XX Query Match: 32.97% Indels: 8
XX DB: 10 Gaps: 3
XX
XX US-10-691-590-1 (1-89) x ADG27463 (1-1949)
XX
XX 1 ArgArgGluGln-----GluGluGluSerGluGluGluThrPheGlyGluPheGln 17
XX
```

```
Db 1413 CGCGCGGAACAAGAGTGGGAAGAGAGGAGGAGATGATGNAAGAAGAGGAGGAGTAACAGA 1472
Qy 18 GlnValLys-----AlaProLeuSerProGlyAspValPheValAlaProAlaGly 34
Db 1473 GAGGTGCGTAGGTACACAGCGAGGTTGAAGGAGGCGATGTGTTTCATCGCCAGCAGCT 1532
Qy 35 HisAlaValThrPhePheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeu 54
Db 1533 CATCCAGTAGCCATCAACGCTTCTCCGAA-----CTCCATCTGCTTGGCTTCGGTATC 1586
Qy 55 AsnAlaGlnAsnAsnGlnArgIlePheLeuAlaGly 66
Db 1587 AACGCTGAATAACAACACACAGATCTTCTTCGAGGT 1622
XX
XX RESULT 13
XX ADL18492
XX ID ADL18492 standard; cDNA; 1722 BP.
XX AC
XX ADL18492;
XX
XX 06-MAY-2004 (first entry)
XX
XX Maize globulin-1 S allele precursor encoding cDNA SEQ ID NO:1.
XX
XX plant; plant molecular biology; commercial; research; molecular marker;
XX plant breeding; maize; globulin-1 S allele precursor; gene; ss.
XX
XX Zea mays.
XX
XX Key Location/Qualifiers
XX CDS 1..1722
XX FT /*tag= a
XX FT /product= "globulin-1 S allele precursor"
XX
XX WO2003027249-A2.
XX
XX 03-APR-2003.
XX
XX 26-SEP-2002; 2002WO-US030475.
XX
XX 26-SEP-2001; 2001US-0325277P.
XX 04-APR-2002; 2002US-0370526P.
XX 04-APR-2002; 2002US-0370620P.
XX
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
XX Su W, Andon N, Haynes P, Briggs SP, Cooper B, Glazebrook J;
XX PI Goff SA, Katagiri F, Kreps J, Moughamer T, Provart N, Ricke D;
XX PI Zhu T;
XX
XX WPI: 2003-354649/33.
XX P-PSDB; ADL18493.
XX
XX New isolated high protein-phenotype-associated plant nucleic acid
XX molecule, useful in plant molecular biology, and specifically for
XX altering protein content or level in plants, and developing molecular
XX markers for plant breeding.
XX
XX Claim 1; SEQ ID NO 1; 163pp; English.
XX
XX The present invention describes an isolated nucleic acid molecule (I)
XX comprising a plant nucleotide sequence or its complement which hybridises
XX under low, moderate or high stringency conditions to a nucleic acid
XX segment encoding a polypeptide comprising any one of SEQ ID NO:1-36. The
XX where the nucleotide sequence does not encode any of SEQ ID NO:1-36. The
XX methods and compositions of the present invention are useful in plant
XX molecular biology, particularly for commercial and research purposes, and
XX more specifically for altering the protein content or level in plants,
XX and to develop molecular markers for plant breeding. The present sequence
XX encodes the maize globulin-1 S allele precursor, which is used in the
XX exemplification of the present invention.
XX
```



PT three dimensional coordinates, three dimensional structured and models  
PT and their uses.

XX PS Disclosure; Page 1285-1287; 1298pp; Japanese.  
XX CC The present invention relates to a new Glycinin characterised by the  
CC atomic coordinate data fully defined in the specification. The structure  
CC can be used for improving processability of soya protein. The present  
CC nucleic acid sequence encodes the Glycine max (soybean) var. Wassuzunari  
CC protein #3, as described in the specification  
XX  
SQ Sequence 1680 BP; 567 A; 385 C; 395 G; 333 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 2,65e-10 Length: 1680  
Score: 143.00 Matches: 31  
Percent Similarity: 66.67% Conservative: 13  
Best Local Similarity: 46.97% Mismatches: 20  
Query Match: 31.22% Indels: 2  
DB: 6 Gaps: 1

US-10-691-590-1 (1-89) x ABS55197 (1-1680)

Qy 1 ArgArgGluGlnGluGluSerGluGluGluThrPheGlyGluPheGlnGlnVallys 20  
Db 1294 AAAGAACAACAACAGAGGGCAGCAACCTTTGGAAGTGCAGAAATATAGA 1353  
Qy 21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40  
Db 1354 GCTGAATTGTCTGAACAAGATATATTGTATCCAGCAGGTATCCAGTTGGTCAAC 1413  
Qy 41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnGln 60  
Db 1414 GCTACCTCAGAT-----CTGAATTCTTTGTTTGGTATCAATGCCGAGAACACCAG 1467  
Qy 61 ArgIlePheLeuAlaGly 66  
Db 1468 AGGAACCTTCCTTCAGGT 1485

Search completed: November 5, 2005, 03:06:24  
Job time : 578.826 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 5, 2005, 02:34:02 ; Search time 192.704 Seconds  
(without alignments)  
755.710 Million cell updates/sec

Title: US-10-691-590-1

Perfect score: 458

Sequence: 1 RREQEESBEETGFEQQVK.....FLNHKQNTNVIKFTVKASAY 89

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US10691590/runat\_02112005\_120023\_13203/app\_query.fasta\_1.462  
-DB=Issued Patents NA -SUFFIX=tni -MINMATCH=0.1 -LOOPCL=0  
-LOOPT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10691590@cgn2\_1\_93@runat\_02112005\_120023\_13203 -NCPU=6 -ICPU=3  
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:\*

1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PTUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	339	74.0	1867	1	US-07-955-905A-1
2	156.5	34.2	1930	4	US-09-106-872A-3
3	156.5	34.2	2032	4	US-09-106-872A-21
4	156.5	34.2	2032	4	US-09-191-593-5
5	156.5	34.2	2041	4	US-09-106-872A-23
6	151	33.0	1949	4	US-09-106-872A-19
7	151	33.0	1949	4	US-09-191-593-4
8	142	31.0	1920	3	US-09-108-010B-2
9	142	31.0	1920	4	US-09-758-652-2
10	142	31.0	1920	4	US-10-684-651-2
11	140	30.6	1320	3	US-09-108-010B-3
12	140	30.6	1320	4	US-09-758-652-3

13	140	30.6	1320	4	US-10-684-651-3	Sequence 3, Appli
14	137	29.9	1818	3	US-09-108-010B-1	Sequence 1, Appli
15	137	29.9	1818	4	US-09-758-652-1	Sequence 1, Appli
16	137	29.9	1818	4	US-10-684-651-1	Sequence 1, Appli
17	119	26.0	1590	3	US-09-323-195A-7	Sequence 7, Appli
18	114	24.9	1581	3	US-09-323-195A-16	Sequence 16, Appli
19	114	24.9	1596	3	US-09-323-195A-19	Sequence 19, Appli
20	109.5	23.9	1413	3	US-09-323-195A-8	Sequence 8, Appli
21	100	21.8	1924	3	US-09-424-283-5	Sequence 5, Appli
22	71	15.5	116866	4	US-09-949-016-17557	Sequence 17557, A
23	69.5	15.2	2677	4	US-09-949-016-1924	Sequence 1924, Ap
24	69.5	15.2	19237	4	US-09-949-016-13666	Sequence 13666, A
25	68.5	15.0	1182	1	US-07-640-476-4	Sequence 4, Appli
26	68.5	15.0	1185	6	5290690-1	Patent No. 5290690
27	68.5	15.0	1185	6	5290690-1	Patent No. 5290690
28	68.5	15.0	1566	6	5290690-3	Patent No. 5290690
29	68.5	15.0	1566	6	5290690-3	Sequence 1535, Ap
30	67.5	14.7	805	4	US-09-270-767-1535	Sequence 1535, A
31	67.5	14.7	805	4	US-09-270-767-16817	Sequence 16817, A
32	67	14.6	1551	3	US-09-108-010B-15	Sequence 15, Appli
33	67	14.6	1551	4	US-09-758-652-15	Sequence 15, Appli
34	67	14.6	1551	4	US-10-684-651-15	Sequence 15, Appli
35	67	14.6	5066	4	US-09-824-574-1	Sequence 1, Appli
36	67	14.6	1664976	4	US-08-916-421B-1	Sequence 1, Appli
37	67	14.6	1664976	4	US-09-692-570-1	Sequence 386, App
38	65	14.2	1785	4	US-09-614-221A-386	Sequence 1, Appli
39	65	14.2	2103	4	US-08-933-711B-1	Sequence 881, App
40	64.5	14.1	411	4	US-09-107-433-881	Sequence 7236, Ap
41	64.5	14.1	690	4	US-09-252-991A-7236	Sequence 3, Appli
42	64.5	14.1	1300	3	US-08-971-782-3	Sequence 3, Appli
43	64.5	14.1	1300	3	US-09-309-026-3	Sequence 2548, Ap
44	64.5	14.1	1380	4	US-09-583-110-2548	Sequence 8719, Ap
45	64.5	14.1	1638	4	US-09-902-540-8719	

#### ALIGNMENTS

RESULT 1  
US-07-955-905A-1  
; Sequence 1, Application US/07955905A  
; Patent No. 5770433  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: RECOMBINANT 47 AND 31 kD COCOA PROTEINS AND  
; TITLE OF INVENTION: PRECURSOR  
; NUMBER OF SEQUENCES: 28  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/955,905A  
; FILING DATE: 21-JAN-1993  
; CLASSIFICATION: 435  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1867 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; ORIGINAL SOURCE:  
; ORGANISM: Theobroma cacao  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 14..1711  
; US-07-955-905A-1

Alignment Scores:  
Pred. No.: 5,71e-40  
Score: 339.00  
Percent Similarity: 100.00%  
Length: 1867  
Matches: 66  
Conservative: 1







Db 1544 AGGTACACAGCGAGGTTGAAGGAAGCGATGTGTTTCATCATGCCAGCAGCTCATCCAGTA 1603  
Qy 38 ThrPhePheNlaSerIysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGln 57  
Db 1604 GCCATCAACGCTTCCTCCGAA-----CTCCATCTGCTTGGCTTCGGTATCAACGCTGAA 1657  
Qy 58 AsnAsnGlnArgIlePheLeuAlaGly 66  
Db 1658 AACACACACAGAAATCTTCTTGCAGGT 1684

## RESULT 6

US-09-106-872A-19  
; Sequence 19, Application US/09106872A  
; Patent No. 6486311  
; GENERAL INFORMATION:  
; APPLICANT: Burks Jr., A. Wesley  
; APPLICANT: Stanley, J. Steven  
; APPLICANT: Cockrell, Gael  
; APPLICANT: King, Nina B.  
; APPLICANT: Sampson, Hugh A.  
; APPLICANT: Helm, Ricki M.  
; APPLICANT: Bannon, Gary A.  
; TITLE OF INVENTION: Peanut Allergens and Methods  
; FILE REFERENCE: HS 103 CIP  
; CURRENT APPLICATION NUMBER: US/09/106,872A  
; CURRENT FILING DATE: 1999-06-29  
; PRIOR APPLICATION NUMBER: PCT/US96/15222  
; PRIOR FILING DATE: 1996-09-23  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 19  
; TYPE: DNA  
; ORGANISM: Arachis hypogaea  
US-09-106-872A-19

Alignment Scores:  
Pred. No.: 3.9e-12 Length: 1949  
Score: 151.00 Matches: 36  
Percent Similarity: 63.89% Conservative: 10  
Best Local Similarity: 50.00% Mismatches: 18  
Query Match: 32.97% Indels: 8  
DB: 4 Gaps: 3

US-10-691-590-1 (1-89) x US-09-106-872A-19 (1-1949)

Qy 1 ArgArgGluGln-----GluGluGluSerGluGluGluThrPheGlyGluPheGln 17  
Db 1413 CGCGGGACACAGAGTGGGAGAGAGAGAGAGATGAAGAGAGGAGGAGGAGTAACAGA 1472  
Qy 18 GlnValLys-----AlaProLeuSerProGlyAspValPheValAlaProAlaGly 34  
Db 1473 GAGGTGCGTAGGTACACAGCGAGGTTGAAGGAAGCGCATGTGTTTCATCATGCCAGCAGCT 1532  
Qy 35 HisAlaValThrPhePheAlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeu 54  
Db 1533 CATCCAGTAGCCCATCAACGCTTCCTCCGAA-----CTCATCTGCTTGGCTTCGGTATC 1586  
Qy 55 AsnAlaGlnAsnGlnArgIlePheLeuAlaGly 66  
Db 1587 AACGCTGAACAAACACACAGAAATCTTCTTGCAGGT 1622

## RESULT 7

US-09-191-593-4  
; Sequence 4, Application US/09191593  
; Patent No. 6835824  
; GENERAL INFORMATION:  
; APPLICANT: BURKS, A Wesley, HELM, Ricki M,  
; APPLICANT: COCKRELL, Gael, STANLEY, J Steven,  
; APPLICANT: BANNON, Gary A  
; TITLE OF INVENTION: PEANUT ALLERGENS AND  
; TITLE OF INVENTION: METHODS  
; NUMBER OF SEQUENCES: 67

; CORRESPONDENCE ADDRESS:  
; ADDRESSES: Head, Johnson & Kachigian  
; STREET: 112 W. Center St., Suite 230  
; CITY: Fayetteville  
; STATE: Arkansas AR  
; COUNTRY: United States of America  
; ZIP: 72701  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb  
; MEDIUM TYPE: Storage  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: MS-DOS 6.2  
; SOFTWARE: Wordperfect 6.0C  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/191,593  
; FILING DATE: 13 NOVEMBER 1998  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/717,933  
; FILING DATE: 23 SEPTEMBER 1996  
; APPLICATION NUMBER: US 07/998,377  
; FILING DATE: 30 DECEMBER 1992  
; APPLICATION NUMBER: US 08/158,704  
; FILING DATE: 29 NOVEMBER 1993  
; APPLICATION NUMBER: US 60/009,455  
; FILING DATE: 29 DECEMBER 1995  
; APPLICATION NUMBER: US 08/610,424  
; FILING DATE: 04 MARCH 1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: ALEXANDER, DANIEL R  
; REGISTRATION NUMBER: 32,604  
; REFERENCE/DOCKET NUMBER: ARK00895601B  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (501) 582-9111  
; TELEFAX: (501) 521-4931  
; TELEX: No. 6835824 applicable  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1949 bases  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA  
; DESCRIPTION: identified as Ara h I Alpha P17  
; HYPOTHEetical: No  
; ANTI-SENSE: No  
; FRAGMENT TYPE: No. 6835824 applicable  
; ORIGINAL SOURCE:  
; ORGANISM: Arachis hypogaea  
; STRAIN: Florunner  
; INDIVIDUAL ISOLATE: Clone P17  
; DEVELOPMENTAL STAGE: Seed  
; HAPLOTYPE: No. 6835824 applicable  
; TISSUE TYPE: Seed mRNA, cDNA library  
; CELL TYPE: No. 6835824 applicable  
; CELL LINE: No. 6835824 applicable  
; ORGANELLE: No. 6835824 applicable  
; IMMEDIATE SOURCE:  
; LIBRARY: Florunner seed cDNA expression  
; LIBRARY: library in Uni-ZAP XR vector  
; CLONE: P17  
; POSITION IN GENOME:  
; CHROMOSOME/SEGMENT: No. 6835824 applicable  
; MAP POSITION: No. 6835824 applicable  
; UNITS: No. 6835824 applicable  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 3..1847  
; IDENTIFICATION METHOD: By agreement with  
; IDENTIFICATION METHOD: protein information and established  
; IDENTIFICATION METHOD: consensus sequence  
; OTHER INFORMATION: Seed storage protein and  
; OTHER INFORMATION: allergen



```

; NAME: LYNNE M. CHRISTENBURY
; REGISTRATION NUMBER: 30,971
; REFERENCE/DOCKET NUMBER: BB-1071-A
; TELEPHONE: 302-992-5481
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1920 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-684-651-2

Alignment Scores:
Pred. No.:      8,19e-11      Length:      1920
Score:          142.00        Matches:     31
Percent Similarity: 66.67%    Conservative: 13
Best Local Similarity: 46.97% Mismatches:    20
Query Match:     31.00%      Indels:       2
DB:              4           Gaps:         1

US-10-691-590-1 (1-89) x US-09-758-652-2 (1-1920)
Qy   1 ArgArgGluGlnGluSerGluGluThrPheGlyGluPheGlnValLys 20
Db   1534 AAAGAACAACAGCAGGCAGCAACAGGAAGCAACCTTTTGGAAATCGCGAATAATAGA 1593
Qy   21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
Db   1594 GCTGAATTGTCTGAACAAGATATATTGTGAATCCAGCAGGTTATCCAGTTATCGTCAAC 1653
Qy   41 AlaSerLyAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnGln 60
Db   1654 GCTACTTCAGAT-----CTGAATTTCTTTTGTTATCATATGCCGAGAACACACAG 1707
Qy   61 ArgilePheLeuAlaGly 66
Db   1708 AGGAACCTCTTCGAGGT 1725

RESULT 10
US-10-684-651-2
; Sequence 2, Application US/10684651
; Patent No. 6828491
; GENERAL INFORMATION:
; APPLICANT: ANTHONY JOHN KINNEY
; GARY MICHAEL FADER
; TITLE OF INVENTION: SUPPRESSION OF SPECIFIC CLASSES OF SOYBEAN SEED PROTEIN GENES
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 95
; SOFTWARE: MICROSOFT WORD 7.0A
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/684,651
; FILING DATE: 14-Oct-2003
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/758,652
; FILING DATE: 11-Jan-2001
; APPLICATION NUMBER: 60/019,940
; FILING DATE: JUNE 14,1996
; ATTORNEY/AGENT INFORMATION:

; REFERENCE/DOCKET NUMBER: BB-1071-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-992-5481
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1920 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-758-652-2

Alignment Scores:
Pred. No.:      8,19e-11      Length:      1920
Score:          142.00        Matches:     31
Percent Similarity: 66.67%    Conservative: 13
Best Local Similarity: 46.97% Mismatches:    20
Query Match:     31.00%      Indels:       2
DB:              4           Gaps:         1

US-10-691-590-1 (1-89) x US-09-758-652-2 (1-1920)
Qy   1 ArgArgGluGlnGluSerGluGluThrPheGlyGluPheGlnValLys 20
Db   1534 AAAGAACAACAGCAGGCAGCAACAGGAAGCAACCTTTTGGAAATCGCGAATAATAGA 1593
Qy   21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
Db   1594 GCTGAATTGTCTGAACAAGATATATTGTGAATCCAGCAGGTTATCCAGTTATCGTCAAC 1653
Qy   41 AlaSerLyAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnGln 60
Db   1654 GCTACTTCAGAT-----CTGAATTTCTTTTGTTATCATATGCCGAGAACACACAG 1707
Qy   61 ArgilePheLeuAlaGly 66
Db   1708 AGGAACCTCTTCGAGGT 1725

RESULT 11
US-09-108-010B-3
; Sequence 3, Application US/09108010B
; Patent No. 6362399
; GENERAL INFORMATION:
; APPLICANT: ANTHONY JOHN KINNEY
; GARY MICHAEL FADER
; TITLE OF INVENTION: SUPPRESSION OF SPECIFIC CLASSES OF SOYBEAN SEED PROTEIN GENES
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 95
; SOFTWARE: MICROSOFT WORD 7.0A
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/108,010B
; FILING DATE: 30-Jun-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/019,940
; FILING DATE: JUNE 14,1996
; ATTORNEY/AGENT INFORMATION:

```





ZIP: 19898  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.50 INCH  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: MICROSOFT WINDOWS 95  
SOFTWARE: MICROSOFT WORD 7.0A  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/758,652  
FILING DATE: 11-Jan-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/019,940  
FILING DATE: JUNE 14,1996  
ATTORNEY/AGENT INFORMATION:  
NAME: LYNNE M. CHRISTENBURY  
REGISTRATION NUMBER: 30,971  
REFERENCE/DOCKET NUMBER: BB-1071-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 302-992-5481  
TELEFAX: 302-773-0164  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1818 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-758-652-1

Alignment Scores:  
Pred. No.: 4.16e-10 Length: 1818  
Score: 137.00 Matches: 30  
Percent Similarity: 66.67% Conservative: 14  
Best Local Similarity: 45.45% Mismatches: 20  
Query Match: 29.91% Indels: 2  
DB: 4 Gaps: 1

US-10-691-590-1 (1-89) x US-09-758-652-1 (1-1818)

Qy 1 ArgArgGluGlnGluGluSerGluGluThrPheGlyGluPheGlnGlnValLys 20  
Db 1432 AAAGAACAAACAGGAGCAGCAACCTTTGGAAAGTCCGAAATATAGA 1491  
Qy 21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40  
Db 1492 GCCGAATTGTCTGAACAGATATATTTGTAATCCGACGAGTTATCCAGTTGTGGTCAAC 1551  
Qy 41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGln 60  
Db 1552 GCTACCTCAAT-----CTGAATTCTTTGCTATTGTTATTATGCGAGAACACACAG 1605  
Qy 61 ArgIlePheLeuAlaGly 66  
Db 1606 AGGAACCTCTCGCAGGT 1623

Search completed: November 5, 2005, 06:01:14  
Job time : 201.704 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 5, 2005, 01:54:58 ; Search time 3520.53 Seconds  
(without alignments)  
962.275 Million cell updates/sec

Title: US-10-691-590-1

Perfect score: 458

Sequence: 1 RREQESESEETFGFEQQVK.....FLNHKQNTNVKFTVKASAY 89

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US10691590/runat\_02112005\_120022\_13171/app\_query.fasta\_1.462  
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10691590@cgn\_1\_1\_4385@runa\_02112005\_120022\_13171 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:\*

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_hic:\*  
4: gb\_est3:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_gss1:\*  
9: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	339	74.0	338	CA795230	CA795230 Cac BL 22
2	339	74.0	359	CA796844	CA796844 Cac BL 38
3	339	74.0	487	CA794382	CA794382 Cac BL 16
4	339	74.0	498	CA794271	CA794271 Cac BL 17
5	339	74.0	498	CA794288	CA794288 Cac BL 17
6	339	74.0	499	CA794282	CA794282 Cac BL 17
7	339	74.0	500	CA797354	CA797354 Cac BL 44
8	339	74.0	505	CA795194	CA795194 Cac BL 22
9	339	74.0	505	CA796872	CA796872 Cac BL 39

10	339	74.0	508	6	CA795155	CA795155 Cac BL 21
11	339	74.0	510	6	CA795157	CA795157 Cac BL 21
12	339	74.0	510	6	CA797074	CA797074 Cac BL 41
13	339	74.0	510	6	CA797130	CA797130 Cac BL 42
14	339	74.0	511	6	CA795185	CA795185 Cac BL 22
15	339	74.0	512	6	CA797107	CA797107 Cac BL 41
16	339	74.0	514	6	CA796326	CA796326 Cac BL 33
17	339	74.0	517	6	CA795470	CA795470 Cac BL 25
18	339	74.0	517	6	CA797325	CA797325 Cac BL 44
19	339	74.0	524	6	CA794505	CA794505 Cac BL 15
20	339	74.0	524	6	CA796851	CA796851 Cac BL 38
21	339	74.0	526	6	CA795003	CA795003 Cac BL 20
22	339	74.0	526	6	CA795423	CA795423 Cac BL 24
23	339	74.0	528	6	CA794978	CA794978 Cac BL 20
24	339	74.0	528	6	CA796838	CA796838 Cac BL 38
25	339	74.0	528	6	CA796893	CA796893 Cac BL 39
26	339	74.0	529	6	CA794376	CA794376 Cac BL 16
27	339	74.0	530	6	CA796006	CA796006 Cac BL 30
28	339	74.0	530	6	CA797179	CA797179 Cac BL 42
29	339	74.0	533	6	CA794366	CA794366 Cac BL 16
30	339	74.0	534	6	CA796817	CA796817 Cac BL 38
31	339	74.0	534	6	CA797168	CA797168 Cac BL 42
32	339	74.0	536	6	CA794760	CA794760 Cac BL 18
33	339	74.0	536	6	CA796191	CA796191 Cac BL 32
34	339	74.0	537	6	CA794789	CA794789 Cac BL 18
35	339	74.0	537	6	CA797261	CA797261 Cac BL 43
36	339	74.0	537	6	CA797571	CA797571 Cac BL 46
37	339	74.0	538	6	CA797162	CA797162 Cac BL 42
38	339	74.0	539	6	CA794360	CA794360 Cac BL 16
39	339	74.0	539	6	CA794367	CA794367 Cac BL 16
40	339	74.0	539	6	CA794934	CA794934 Cac BL 19
41	339	74.0	540	6	CA797170	CA797170 Cac BL 42
42	339	74.0	540	6	CA797603	CA797603 Cac BL 46
43	339	74.0	541	6	CA796465	CA796465 Cac BL 34
44	339	74.0	541	6	CA797248	CA797248 Cac BL 43
45	339	74.0	544	6	CA796176	CA796176 Cac BL 31

#### ALIGNMENTS

RESULT 1  
CA795230 338 bp mRNA linear EST 05-DEC-2002  
LOCUS Cac BL 226 Cac BL (Bean and Leaf from Amelonado type Cacao)  
DEFINITION Theobroma cacao cDNA clone Cac\_BL\_226 5', mRNA sequence.  
ACCESSION CA795230 GI:26052306  
VERSION CA795230  
KEYWORDS EST.  
SOURCE Theobroma cacao (cacao)  
ORGANISM Theobroma cacao  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae; Theobroma.

#### REFERENCE

1 (bases 1 to 338)

#### AUTHORS

Jones, P.G., Allaway, D., Gilmour, D.M., Harris, C., Rankin, D., Retzel, E.R., and Jones, C.A.

#### TITLE

Gene discovery and microarray analysis of cacao (Theobroma cacao L.) varieties

#### JOURNAL

Planta 216 (2), 255-264 (2002)

#### MEDLINE

22337596

#### PUBMED

12447539

#### COMMENT

Contact: Jones, Paul

Masterfoods

3d Dundee Road, Slough, Berkshire, UK, SL1 4LG

Tel: +44 1664 416644

Email: Paul.Jones@eu.affem.com

Seq primer: T3.

Location/Qualifiers

1..338

/organism="Theobroma cacao"

/mol\_type="mRNA"

/strain="Amelonado type"



```
/dev_stage="maturity"  
/lab_host="XL-1 Blue MRF"  
/clone_lib="Cac_BL (Bean and Leaf from Amelonardo type  
Cacao)"  
/note="Vector: pBK-CMV; Bean and leaf tissue from an  
Amelonado type Cacao tree."
```

## ORIGIN

```
Alignment Scores:  
Pred. No.: 6.12e-36 Length: 497  
Score: 339.00 Matches: 66  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 98.51% Mismatches: 0  
Query Match: 74.02% Indels: 0  
DB: 6 Gaps: 0
```

US-10-691-590-1 (1-89) x CA794382 (1-487)

```
QY 1 ArgArgGluGlnGluGluSerGluGluThrPheGlyGluPheGlnGlnVallys 20  
|||||  
Db 170 AGAAGAGACACAGAGAGAGTCAAGAGAGGACATTTGGAGAAATTCAGCAGGTCAA 229  
|||||  
QY 21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40  
|||||  
Db 230 GCCCCATTGTCACTGTGTGACGCTTTGTAGCCCGCAGGCCATGCGATTACATTTCTTT 289  
|||||  
QY 41 AlaSerIysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAenGln 60  
|||||  
Db 290 GCATCCAAAGACAGCCCTGAATGCGTTTGGACTCAACGCCAGAACCAACCAG 349  
|||||  
QY 61 ArgIlePheLeuAlaGlyArg 67  
|||||  
Db 350 AGAATTTTCCTTCGAGGGAAA 370  
|||||
```

## RESULT 4

CA794271

LOCUS

CA794271  
Cac BL 1751 Cac BL (Bean and Leaf from Amelonardo type Cacao)  
Theobroma cacao cDNA clone Cac\_BL\_1751 5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

```
EST.  
Theobroma cacao (cacao)  
Theobroma cacao  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;  
Theobroma.
```

REFERENCE

AUTHORS

Jones, P.G., Allaway, D., Gilmour, D.M., Harris, C., Rankin, D.,  
Retzel, E.R. and Jones, C.A.

TITLE

Gene discovery and microarray analysis of cacao (Theobroma cacao

L.) varieties

JOURNAL

MEDLINE

PUBMED

COMMENT

Planta 216 (2), 255-264 (2002)  
12447539  
Contact: Jones, Paul  
Masterfoods  
3d Dundee Road, Slough, Berkshire, UK, SL1 4LG  
Tel: +44 1664 416644  
Email: Paul.Jones@eu.affem.com

Seq primer: T3.

Location/Qualifiers

1. .498

/organism="Theobroma cacao"

/mol\_type="mRNA"

/strain="Amelonado type"

/db\_xref="taxon:3641"

/clone="Cac BL 1751"

/tissue type="Mature leaf and mature bean"

/cell type="Whole organ"

/dev\_stage="maturity"

/lab\_host="XL-1 Blue MRF"

FEATURES

source

1. .498

/organism="Theobroma cacao"

/mol\_type="mRNA"

/strain="Amelonado type"

/db\_xref="taxon:3641"

/clone="Cac BL 1751"

/tissue type="Mature leaf and mature bean"

/cell type="Whole organ"

/dev\_stage="maturity"

/lab\_host="XL-1 Blue MRF"

```
/clone_lib="Cac_BL (Bean and Leaf from Amelonardo type  
Cacao)"  
/note="Vector: pBK-CMV; Bean and leaf tissue from an  
Amelonado type Cacao tree."
```

## ORIGIN

```
Alignment Scores:  
Pred. No.: 6.3e-36 Length: 498  
Score: 339.00 Matches: 66  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 98.51% Mismatches: 0  
Query Match: 74.02% Indels: 0  
DB: 6 Gaps: 0
```

US-10-691-590-1 (1-89) x CA794271 (1-498)

```
QY 1 ArgArgGluGlnGluGluSerGluGluThrPheGlyGluPheGlnGlnVallys 20  
|||||  
Db 185 AGAAGAGAACACAGAGAGAGTCAAGAGAGGAGACATTTGGAGAAATTCAGCAGGTCAA 244  
|||||  
QY 21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40  
|||||  
Db 245 GCCCCATTGTCACTGTGTGACGCTTTGTAGCCCGCAGGCCATGCGATTACATTTCTTT 304  
|||||  
QY 41 AlaSerIysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAenGln 60  
|||||  
Db 305 GCATCCAAAGACAGCCCTGAATGCGTTTGGACTCAACGCCAGAACCAACCAG 364  
|||||  
QY 61 ArgIlePheLeuAlaGlyArg 67  
|||||  
Db 365 AGAATTTTCCTTCGAGGGAAA 385  
|||||
```

## RESULT 5

CA794288

LOCUS

DEFINITION

Cac BL 1724 Cac BL (Bean and Leaf from Amelonardo type Cacao)  
Theobroma cacao cDNA clone Cac\_BL\_1724 5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

```
EST.  
Theobroma cacao (cacao)  
Theobroma cacao  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;  
Theobroma.
```

REFERENCE

AUTHORS

Jones, P.G., Allaway, D., Gilmour, D.M., Harris, C., Rankin, D.,  
Retzel, E.R. and Jones, C.A.

TITLE

Gene discovery and microarray analysis of cacao (Theobroma cacao

L.) varieties

JOURNAL

MEDLINE

PUBMED

COMMENT

Planta 216 (2), 255-264 (2002)  
2237596  
12447539  
Contact: Jones, Paul  
Masterfoods  
3d Dundee Road, Slough, Berkshire, UK, SL1 4LG  
Tel: +44 1664 416644  
Email: Paul.Jones@eu.affem.com

Seq primer: T3.

Location/Qualifiers

1. .498

/organism="Theobroma cacao"

/mol\_type="mRNA"

/strain="Amelonado type"

/db\_xref="taxon:3641"

/clone="Cac BL 1724"

/tissue type="Mature leaf and mature bean"

/cell type="Whole organ"

/dev\_stage="maturity"

/lab\_host="XL-1 Blue MRF"

/clone\_lib="Cac\_BL (Bean and Leaf from Amelonardo type

Cacao)"



```

Alignment Scores:
Pred. No.: 6,33e-36 Length: 500
Score: 339.00 Matches: 66
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 98.51% Indels: 0
DB: 74.02% Gaps: 0
DB: 6

US-10-691-590-1 (1-89) x CA797354 (1-500)

QY 1 ArgArgGluGluGluGluSerGluGluGluThrPheGlyGluPheGluGluVallys 20
Db 110 AGAAGAAACAGAGAAGAGTGCAGAGAGAGAGACATTTGGAGAAATTCAGCAGGTCAAA 169
QY 21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
Db 170 GCCCATTTGCACCTGGTGCAGTCTTTGTAGCCCCGCGAGCCATGCAGTTACATTTCTTT 229
QY 41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnGln 60
Db 230 GCATCCAAAGACAGCCCTGAATGCAGTTGCGTTGGACTCAACGCCAGAACACACAG 289
QY 61 ArgIlePheLeuAlaGlyArg 67
Db 290 AGAATTTCTTCGAGGAAA 310

RESULT 8
CA795194 505 bp mRNA linear EST 05-DEC-2002
LOCUS Cac BL 2226 Cac BL (Bean and Leaf from Amelonardo type Cacao)
DEFINITION Theobroma cacao cDNA clone Cac BL 2226 5', mRNA sequence.
ACCESSION CA795194
VERSION CA795194.1 GI:26052270
KEYWORDS EST.
SOURCE Theobroma cacao (cacao)
ORGANISM Theobroma cacao
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;
Theobroma.
REFERENCE 1 (bases 1 to 505)
AUTHORS Jones,P.G., Allaway,D., Gilmour,D.M., Harris,C., Rankin,D.,
Retzel,E.R. and Jones,C.A.
TITLE Gene discovery and microarray analysis of cacao (Theobroma cacao
L.) varieties
JOURNAL Planta 216 (2), 255-264 (2002)
MEDLINE 22337596
PUBMED 12447539
COMMENT Contact: Jones, Paul
Masterfoods
3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
Tel: +44 1664 416644
Email: Paul.Jones@eu.affem.com
Seq primer: T3.
FEATURES
source
1..505
/organism="Theobroma cacao"
/mol_type="mRNA"
/strain="Amelonado type"
/db_xref="taxon:3641"
/clone="Cac BL 2226"
/tissue_type="Mature leaf and mature bean"
/cell_type="Whole organ"
/dev_stage="maturity"
/lab_host="XL-1 Blue MRF"
/clone_lib="Cac BL (Bean and Leaf from Amelonardo type
Cacao)"
/note="Vector: pBK-CMV; Bean and leaf tissue from an
Amelonardo type Cacao tree."

ORIGIN
Alignment Scores:
Pred. No.: 6,41e-36 Length: 505
Score: 339.00 Matches: 66
Percent Similarity: 100.00% Conservative: 1

US-10-691-590-1 (1-89) x CA795194 (1-505)

QY 1 ArgArgGluGluGluGluSerGluGluGluThrPheGlyGluPheGluGluVallys 20
Db 62 AGAAGAAACAGAGAAGAGTGCAGAGAGAGAGACATTTGGAGAAATTCAGCAGGTCAAA 121
QY 21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
Db 122 GCCCATTTGCACCTGGTGCAGTCTTTGTAGCCCCGCGAGCCATGCAGTTACATTTCTTT 181
QY 41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnGln 60
Db 182 GCATCCAAAGACAGCCCTGAATGCAGTTGCGTTGGACTCAACGCCAGAACACACAG 241
QY 61 ArgIlePheLeuAlaGlyArg 67
Db 242 AGAATTTCTTCGAGGAAA 262

RESULT 9
CA796872 505 bp mRNA linear EST 05-DEC-2002
LOCUS Cac BL 3924 Cac BL (Bean and Leaf from Amelonardo type Cacao)
DEFINITION Theobroma cacao cDNA clone Cac BL 3924 5', mRNA sequence.
ACCESSION CA796872
VERSION CA796872.1 GI:26053958
KEYWORDS EST.
SOURCE Theobroma cacao (cacao)
ORGANISM Theobroma cacao
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;
Theobroma.
REFERENCE 1 (bases 1 to 505)
AUTHORS Jones,P.G., Allaway,D., Gilmour,D.M., Harris,C., Rankin,D.,
Retzel,E.R. and Jones,C.A.
TITLE Gene discovery and microarray analysis of cacao (Theobroma cacao
L.) varieties
JOURNAL Planta 216 (2), 255-264 (2002)
MEDLINE 22337596
PUBMED 12447539
COMMENT Contact: Jones, Paul
Masterfoods
3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
Tel: +44 1664 416644
Email: Paul.Jones@eu.affem.com
Seq primer: T3.
FEATURES
source
1..505
/organism="Theobroma cacao"
/mol_type="mRNA"
/strain="Amelonado type"
/db_xref="taxon:3641"
/clone="Cac BL 3924"
/tissue_type="Mature leaf and mature bean"
/cell_type="Whole organ"
/dev_stage="maturity"
/lab_host="XL-1 Blue MRF"
/clone_lib="Cac BL (Bean and Leaf from Amelonardo type
Cacao)"
/note="Vector: pBK-CMV; Bean and leaf tissue from an
Amelonardo type Cacao tree."

ORIGIN
Alignment Scores:
Pred. No.: 6,41e-36 Length: 505
Score: 339.00 Matches: 66
Percent Similarity: 100.00% Conservative: 1

```

Best Local Similarity:	98.51%	Mismatches:	0
Query Match:	74.02%	Indels:	0
DB:	6	Gaps:	0
US-10-691-590-1 (1-89) x CA796872 (1-505)			
QY	1	ArgArgGluGlnGluGluSerGluGluThrPheGlyGluPheGlnValValys	20
Db	41	AGAAGAACACAGAGAGAGTCAAGAGAGGACATTTGGAGATTCACAGAGTCAAA	100
QY	21	AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe	40
Db	101	GCCTCATTTGTCACCTGGTGCAGCTTTGTAGCCCGCGCAGGCATGCAGTTACATTC	160
QY	41	AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnGln	60
Db	161	GCATCAAAAGACAGCCCTGAATGCAGTTGCGTTTGGACTCAACGCCCAAGAACACCAG	220
QY	61	ArgIlePheLeuAlaGlyArg	67
Db	221	AGAATTTTCTTCGAGGAAA	241
RESULT 10			
CA795155			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
COMMENT			
FEATURES			
source			
ORIGIN			
Alignment Scores:			
Pred. No.:			
Score:			
Percent Similarity:			
Best Local Similarity:			
Query Match:			
DB:			
Length:			
Matches:			
Conservative:			
Indels:			
Gaps:			

US-10-691-590-1 (1-89) x CA795157 (1-510)

```

QY      1 ArgArgGluGlnGluSerGluGluThrPheGlyGluPheGlnVallys 20
      |||
Db      72 AGAAGAGAACAAAGAGAGT CAGAGAGGACATTTGGAGAATTCACGAGGTCAAA 131
      |||
QY      21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
      |||
Db      132 GCCCATTTGTCACCTGGTGACGCTTTGTAGCCCCCGCAGGCCATCGAGTTACATTTCTTT 191
      |||
QY      41 AlaSerLyAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnGln 60
      |||
Db      192 GCATCCAAAGACAGCCCTGAATGCAGTTGCGTTTGGACTCAACGCCCAAGAACACACAG 251
      |||
QY      61 ArgIlePheLeuAlaGlyArg 67
      |||
Db      252 AGAATTTTCTTCGAGGAAA 272
      |||

```

RESULT 12  
CA797074  
LOCUS  
DEFINITION  
Cac\_BU\_4159 Cac\_BL (Bean and Leaf from Amelonardo type Cacao)  
Theobroma cacao cDNA clone Cac\_BU\_4159 5', mRNA sequence.  
CA797074  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

CA797074.1 GI:26054160  
EST.  
Theobroma cacao (cacao)  
Theobroma cacao  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;  
Theobroma.

REFERENCE 1 (bases 1 to 510)  
JONES, P.G., ALLAWAY, D., GILMOUR, D.M., HARRIS, C., RANKIN, D.,  
RETZEL, E.R. and JONES, C.A.  
Gene discovery and microarray analysis of cacao (Theobroma cacao  
L.) varieties  
Planta 216 (2), 255-264 (2002)

JOURNAL  
MEDLINE  
PUBMED  
COMMENT  
Masterfoods  
Contact: Jones, Paul  
3d Dundee Road, Slough, Berkshire, UK, SL1 4LG  
Tel: +44 1664 416644  
Email: Paul.Jones@eu.affem.com  
Seq primer: T3.

FEATURES  
source  
Location/Qualifiers

```

1..510
/organism="Theobroma cacao"
/mol_type="mRNA"
/strain="Amelonado type"
/db_xref="taxon:3641"
/clone="Cac_BU_4159"
/tissue_type="Mature leaf and mature bean"
/cell_type="Whole organ"
/dev_stage="maturity"
/lab_host="XL-1 Blue MRF"
/clone_lib="Cac_BL (Bean and Leaf from Amelonardo type Cacao)"
/notes="Vector: pBK-CMV; Bean and leaf tissue from an Amelonado type Cacao tree."

```

ORIGIN

Alignment Scores:  
Pred. No.: 6.5e-36 Length: 510  
Score: 339.00 Matches: 66  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 98.51% Mismatches: 0  
Query Match: 74.02% Indels: 0  
DB: 6 Gaps: 0

US-10-691-590-1 (1-89) x CA797074 (1-510)

```

QY      1 ArgArgGluGlnGluSerGluGluThrPheGlyGluPheGlnVallys 20
      |||
Db      306 AGAAGAGAACAAAGAGAGT CAGAGAGGACATTTGGAGAATTCACGAGGTCAAA 365
      |||
QY      21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40
      |||
Db      366 GCCCATTTGTCACCTGGTGACGCTTTGTAGCCCCCGCAGGCCATCGAGTTACATTTCTTT 425
      |||
QY      41 AlaSerLyAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnGln 60
      |||
Db      426 GCATCCAAAGACAGCCCTGAATGCAGTTGCGTTTGGACTCAACGCCCAAGAACACACAG 485
      |||
QY      61 ArgIlePheLeuAlaGlyArg 67
      |||
Db      486 AGAATTTTCTTCGAGGAAA 506
      |||

```

RESULT 13  
CA797130  
LOCUS

DEFINITION  
Cac\_BL\_4215 Cac\_BL (Bean and Leaf from Amelonardo type Cacao)  
Theobroma cacao cDNA clone Cac\_BL\_4215 5', mRNA sequence.  
CA797130  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

CA797130.1 GI:26054216  
EST.  
Theobroma cacao (cacao)  
Theobroma cacao  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;  
Theobroma.

REFERENCE 1 (bases 1 to 510)  
JONES, P.G., ALLAWAY, D., GILMOUR, D.M., HARRIS, C., RANKIN, D.,  
RETZEL, E.R. and JONES, C.A.  
Gene discovery and microarray analysis of cacao (Theobroma cacao  
L.) varieties  
Planta 216 (2), 255-264 (2002)

JOURNAL  
MEDLINE  
PUBMED  
COMMENT  
Masterfoods  
Contact: Jones, Paul  
3d Dundee Road, Slough, Berkshire, UK, SL1 4LG  
Tel: +44 1664 416644  
Email: Paul.Jones@eu.affem.com  
Seq primer: T3.

FEATURES  
source  
Location/Qualifiers

```

1..510
/organism="Theobroma cacao"
/mol_type="mRNA"
/strain="Amelonado type"
/db_xref="taxon:3641"
/clone="Cac_BL_4215"
/tissue_type="Mature leaf and mature bean"
/cell_type="Whole organ"
/dev_stage="maturity"
/lab_host="XL-1 Blue MRF"
/clone_lib="Cac_BL (Bean and Leaf from Amelonardo type Cacao)"
/notes="Vector: pBK-CMV; Bean and leaf tissue from an Amelonado type Cacao tree."

```

ORIGIN

Alignment Scores:  
Pred. No.: 6.5e-36 Length: 510  
Score: 339.00 Matches: 66  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 98.51% Mismatches: 0  
Query Match: 74.02% Indels: 0  
DB: 6 Gaps: 0

US-10-691-590-1 (1-89) x CA797130 (1-510)

QY 1 ArgArgGluGlnGluSerGluGluThrPheGlyGluPheGlnVallys 20  
 |||

Db 282 AGAAGAGAACAGAGAGAGTTCAGAGAGGACACATTGGAGAAATTCAGCAGGTCAAA 341

Qy 21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40  
 |||||  
 Db 342 GCCCATTGTCACCTGGTGACGTCTTTGTAGCCCGGCGAGGCCATGCAGTTACATTTCTTT 401  
 |||||

Qy 41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnGln 60  
 |||||  
 Db 402 GCATCAAGACAGAGCCCTGAATGCAGTTGGCTTTGGACTCAACGCCAGAACCAACCAG 461  
 |||||

Qy 61 ArgIlePheLeuAlaGlyArg 67  
 |||||  
 Db 462 AGAATTTCTTCAGGGGAAA 482  
 |||||

RESULT 14  
 CA795185 511 bp mRNA linear EST 05-DEC-2002  
 LOCUS Cac BL 2218 Cac BL (Bean and Leaf from Amelonardo type Cacao)  
 DEFINITION Theobroma cacao cDNA clone Cac\_BL\_2218 5', mRNA sequence.

ACCESSION CA795185  
 VERSION CA795185.1 GI:26052261  
 SOURCE EST.  
 ORGANISM Theobroma cacao (cacao)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;  
 Theobroma.

REFERENCE 1 (bases 1 to 511)  
 AUTHORS Jones,P.G., Allaway,D., Gilmour,D.M., Harris,C., Rankin,D.,  
 Retzel,E.R. and Jones,C.A.  
 TITLE Gene discovery and microarray analysis of cacao (Theobroma cacao  
 L.) varieties  
 JOURNAL Planta 216 (2), 255-264 (2002)  
 MEDLINE 22337596  
 PUBMED 12447539  
 COMMENT Contact: Jones, Paul  
 Masterfoods  
 3d Dundee Road, Slough, Berkshire, UK, SL1 4LG  
 Tel: +44 1664 416644  
 Email: Paul.Jones@eu.effem.com  
 Seq primer: T3.

FEATURES  
 source  
 1..511  
 Location/Qualifiers  
 /organism="Theobroma cacao"  
 /mol\_type="mRNA"  
 /strain="Amelonado type"  
 /db\_xref="taxon:3641"  
 /clone="Cac BL 2218"  
 /tissue\_type="Mature leaf and mature bean"  
 /cell\_type="Whole organ"  
 /dev\_stage="maturity"  
 /lab\_host="XL-1 Blue MRF"  
 /clone\_lib="Cac BL (Bean and Leaf from Amelonardo type Cacao)"  
 /note="Vector: pBK-CMV; Bean and leaf tissue from an Amelonado type Cacao tree."

ORIGIN  
 Alignment Scores:  
 Pred. No.: 6 51e-36 Length: 511  
 Score: 339.00 Matches: 66  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 98.51% Mismatches: 0  
 Query Match: 74.02% Indels: 0  
 DB: 6 Gaps: 0

US-10-691-590-1 (1-89) x CA795185 (1-511)

Qy 1 ArgArgGluGlnGluSerGluGluThrPheGlyGluPheGlnGlnVallys 20  
 |||||  
 Db 271 AGAAGAGAACAGAGAGAGTTCAGAGAGGACATTTGGAGAAATTCAGCAGGTCAAA 330  
 |||||

Qy 21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40  
 |||||  
 Db 331 GCCCATTGTCACCTGGTGACGTCTTTGTAGCCCGGCGAGGCCATGCAGTTACATTTCTTT 390  
 |||||

Qy 41 AlaSerLysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnGln 60  
 |||||  
 Db 391 GCATCAAGACAGAGCCCTGAATGCAGTTGGCTTTGGACTCAACGCCAGAACCAACCAG 450  
 |||||

Qy 61 ArgIlePheLeuAlaGlyArg 67  
 |||||  
 Db 451 AGAATTTCTTCAGGGGAAA 471  
 |||||

RESULT 15  
 CA797107 512 bp mRNA linear EST 05-DEC-2002  
 LOCUS Cac BL 4193 Cac BL (Bean and Leaf from Amelonardo type Cacao)  
 DEFINITION Theobroma cacao cDNA clone Cac\_BL\_4193 5', mRNA sequence.

ACCESSION CA797107  
 VERSION CA797107.1 GI:26054193  
 KEYWORDS EST.  
 SOURCE Theobroma cacao (cacao)  
 ORGANISM Theobroma cacao  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;  
 Theobroma.

REFERENCE 1 (bases 1 to 512)  
 AUTHORS Jones,P.G., Allaway,D., Gilmour,D.M., Harris,C., Rankin,D.,  
 Retzel,E.R. and Jones,C.A.  
 TITLE Gene discovery and microarray analysis of cacao (Theobroma cacao  
 L.) varieties  
 JOURNAL Planta 216 (2), 255-264 (2002)  
 MEDLINE 22337596  
 PUBMED 12447539  
 COMMENT Contact: Jones, Paul  
 Masterfoods  
 3d Dundee Road, Slough, Berkshire, UK, SL1 4LG  
 Tel: +44 1664 416644  
 Email: Paul.Jones@eu.effem.com  
 Seq primer: T3.

FEATURES  
 source  
 1..512  
 Location/Qualifiers  
 /organism="Theobroma cacao"  
 /mol\_type="mRNA"  
 /strain="Amelonado type"  
 /db\_xref="taxon:3641"  
 /clone="Cac BL 4193"  
 /tissue\_type="Mature leaf and mature bean"  
 /cell\_type="Whole organ"  
 /dev\_stage="maturity"  
 /lab\_host="XL-1 Blue MRF"  
 /clone\_lib="Cac BL (Bean and Leaf from Amelonardo type Cacao)"  
 /note="Vector: pBK-CMV; Bean and leaf tissue from an Amelonado type Cacao tree."

ORIGIN  
 Alignment Scores:  
 Pred. No.: 6 53e-36 Length: 512  
 Score: 339.00 Matches: 66  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 98.51% Mismatches: 0  
 Query Match: 74.02% Indels: 0  
 DB: 6 Gaps: 0

US-10-691-590-1 (1-89) x CA797107 (1-512)

Qy 1 ArgArgGluGlnGluSerGluGluThrPheGlyGluPheGlnGlnVallys 20  
 |||||  
 Db 227 AGAAGAGAACAGAGAGAGTTCAGAGAGGACATTTGGAGAAATTCAGCAGGTCAAA 286  
 |||||

Qy 21 AlaProLeuSerProGlyAspValPheValAlaProAlaGlyHisAlaValThrPhePhe 40  
 |||||



```

Db      287 GCCCATTGTACCTGGTGACGTCCTTTGTAGCCCGGCGGCGCATGCAGTTACATTCTTT 346
Qy      41  AlaSerIysAspGlnProLeuAsnAlaValAlaPheGlyLeuAsnAlaGlnAsnAsnGln 60
Db      347 GCATCCAAAGACGAGCCCTGAATGCAGTTGCGTTTGGACTCAACGCCCGAGAACACCAG 406
Qy      61  ArgIlePheLeuAlaGlyArg 67
Db      407 AGAATTTTCCTTGCAGGGAAA 427

```

Search completed: November 5, 2005, 05:56:53  
 Job time : 3527.53 secs

***This Page Blank (uspto)***

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 5, 2005, 01:54:58 ; Search time 1028.47 Seconds  
(without alignments)  
962.275 Million cell updates/sec

Title: US-10-691-590-3

Perfect score: 122

Sequence: 1 RREQESEBETFGFXQVXKPLXPG 26

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US10691590/runat\_02112005\_120022\_13171/app\_query.fasta\_1.462  
-DBS=EST -QFMT=fastap -SUFFIX=rat -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10691590 @CGN\_1\_1\_4385 @runat\_02112005\_120022\_13171 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCKS=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_hic:\*  
4: gb\_est3:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_ges1:\*  
9: gb\_ges2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	117	95.9	256	6	CA794775 Cac BL 18
2	117	95.9	274	6	CA797731 Cac BL 48
3	117	95.9	338	6	CA795230 Cac BL 22
4	117	95.9	351	6	CA795596 Cac BL 26
5	117	95.9	359	6	CA796844 Cac BL 38
6	117	95.9	362	6	CA796065 Cac BL 30
7	117	95.9	364	6	CA797330 Cac BL 44
8	117	95.9	378	6	CA795334 Cac BL 23
9	117	95.9	423	6	CA796843 Cac BL 38

10 117 95.9 430 6 CA794888  
11 117 95.9 441 6 CA797867  
12 117 95.9 463 6 CA795681  
13 117 95.9 465 6 CA794398  
14 117 95.9 468 6 CA797010  
15 117 95.9 474 6 CA795491  
16 117 95.9 475 6 CA796772  
17 117 95.9 477 6 CA794451  
18 117 95.9 482 6 CA795961  
19 117 95.9 487 6 CA794382  
20 117 95.9 490 6 CA796238  
21 117 95.9 491 6 CA797417  
22 117 95.9 495 6 CA794315  
23 117 95.9 498 6 CA794271  
24 117 95.9 498 6 CA794288  
25 117 95.9 498 6 CA794445  
26 117 95.9 498 6 CA797323  
27 117 95.9 499 6 CA794282  
28 117 95.9 499 6 CA794340  
29 117 95.9 499 6 CA794715  
30 117 95.9 500 6 CA794343  
31 117 95.9 500 6 CA797354  
32 117 95.9 500 6 CA798279  
33 117 95.9 503 6 CA797220  
34 117 95.9 505 6 CA795194  
35 117 95.9 505 6 CA796872  
36 117 95.9 508 6 CA794856  
37 117 95.9 508 6 CA795155  
38 117 95.9 508 6 CA795218  
39 117 95.9 508 6 CA797423  
40 117 95.9 509 6 CA797123  
41 117 95.9 510 6 CA795157  
42 117 95.9 510 6 CA797074  
43 117 95.9 510 6 CA797130  
44 117 95.9 511 6 CA795185  
45 117 95.9 512 6 CA797107

#### ALIGNMENTS

RESULT 1  
CA794775

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

FEATURES

source

location

qualifiers

organism

mol\_type

strain

type

name

accession

version

keywords

source

organism

reference

authors

title

journal

medline

pubmed

comment

CA794888 Cac BL 19  
CA797867 Cac BL 51  
CA795681 Cac BL 27  
CA794398 Cac BL 16  
CA797010 Cac BL 40  
CA795491 Cac BL 25  
CA796772 Cac BL 38  
CA794451 Cac BL 15  
CA795961 Cac BL 29  
CA794382 Cac BL 16  
CA796238 Cac BL 32  
CA797417 Cac BL 45  
CA794315 Cac BL 17  
CA794271 Cac BL 17  
CA794288 Cac BL 17  
CA794445 Cac BL 15  
CA797323 Cac BL 44  
CA794282 Cac BL 17  
CA794340 Cac BL 16  
CA794715 Cac BL 10  
CA794343 Cac BL 16  
CA797354 Cac BL 44  
CA798279 Cac BL 60  
CA797220 Cac BL 43  
CA795194 Cac BL 22  
CA796872 Cac BL 39  
CA794856 Cac BL 19  
CA795155 Cac BL 21  
CA795218 Cac BL 22  
CA797423 Cac BL 45  
CA797123 Cac BL 42  
CA795157 Cac BL 21  
CA797074 Cac BL 41  
CA797130 Cac BL 42  
CA795185 Cac BL 22  
CA797107 Cac BL 41

CA794775 256 bp mRNA linear EST 05-DEC-2002  
Cac BL 1827 Cac BL (Bean and Leaf from Amelonado type Cacao)  
Theobroma cacao cDNA clone Cac BL 1827 5', mRNA sequence.

CA794775.1 GI:26051851

EST.

Theobroma cacao (cacao)

Theobroma cacao

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;  
Theobroma.

1 (bases 1 to 256)

Jones, P.G., Allaway, D., Gilmour, D.M., Harris, C., Rankin, D.,  
Retzel, E.R. and Jones, C.A.

Gene discovery and microarray analysis of cacao (Theobroma cacao  
L.) varieties

Planta 216 (2), 255-264 (2002)

22337596

12447539

Contact: Jones, Paul

Masterfoods

3d Dundee Road, Slough, Berkshire, UK, SL1 4LG

Tel: +44 1664 416644

Email: Paul.Jones@eu.affm.com

Seq primer: T3.

Location/Qualifiers

1..256

/organism="Theobroma cacao"

/mol\_type="mRNA"

/strain="Amelonado type"

```

/db_xref="taxon:3641"
/clone="Cac_BL_1827"
/tissue_type="Mature leaf and mature bean"
/cell_type="Whole organ"
/dev_stage="maturity"
/lab_host="XL-1 Blue MRF"
/clone_lib="Cac_BL (Bean and Leaf from Amelonardo type
Cacao)"
/note="Vector: pBK-CMV; Bean and leaf tissue from an
Amelonado type Cacao tree."

ORIGIN
Alignment Scores:
Pred. No.: 3.13e-09 Length: 256
Score: 117.00 Matches: 23
Percent Similarity: 88.46% Conservative: 0
Best Local Similarity: 88.46% Mismatches: 3
Query Match: 95.90% Indels: 0
DB: 6 Gaps: 0

US-10-691-590-3 (1-26) x CA794775 (1-256)

QY 1 ArgAGGluGluGluSerGluGluThrPheGlyGluPhe***GlnVal*** 20
Db 105 AGAAGAGACACAGAGAGAGTCAGAGAGGAGACATTTGGAGATTCCAGCAGGTCAA 164

QY 21 AlaProLeu***ProGly 26
Db 165 GCCCATTTGCACCTGGT 182

RESULT 2
CA797731
LOCUS
DEFINITION
Cac BL 483 Cac BL (Bean and Leaf from Amelonardo type Cacao)
EST 05-DEC-2002
Theobroma cacao cDNA clone Cac_BL_483 5', mRNA sequence.
ACCESSION
CA797731
VERSION
CA797731.1 GI:26054817
KEYWORDS
SOURCE
Theobroma cacao (cacao)
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;
Theobroma.
REFERENCE
1 (bases 1 to 274)
Jones,P.G., Allaway,D., Gilmour,D.M., Harris,C., Rankin,D.,
Retzel,E.R. and Jones,C.A.
Gene discovery and microarray analysis of cacao (Theobroma cacao
L.) varieties
Planta 216 (2), 255-264 (2002)
22337596
12447539
PUBMED
COMMENT
Contact: Jones, Paul
Masterfoods
3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
Tel: +44 1664 416644
Email: Paul.Jones@eu.affem.com
Seq primer: T3.
FEATURES
source
1..274
/organism="Theobroma cacao"
/mol_type="mRNA"
/strain="Amelonado type"
/db_xref="taxon:3641"
/clone="Cac BL 483"
/tissue_type="Mature leaf and mature bean"
/cell_type="Whole organ"
/dev_stage="maturity"
/lab_host="XL-1 Blue MRF"
/clone_lib="Cac_BL (Bean and Leaf from Amelonardo type
Cacao)"
/note="Vector: pBK-CMV; Bean and leaf tissue from an
Amelonado type Cacao tree."

```

```

ORIGIN
Alignment Scores:
Pred. No.: 3.38e-09 Length: 274
Score: 117.00 Matches: 23
Percent Similarity: 88.46% Conservative: 0
Best Local Similarity: 88.46% Mismatches: 3
Query Match: 95.90% Indels: 0
DB: 6 Gaps: 0

US-10-691-590-3 (1-26) x CA797731 (1-274)

QY 1 ArgAGGluGluGluSerGluGluThrPheGlyGluPhe***GlnVal*** 20
Db 63 AGAAGAGACACAGAGAGTCAGAGAGGAGACATTTGGAGATTCCAGCAGGTCAA 122

QY 21 AlaProLeu***ProGly 26
Db 123 GCCCATTTGCACCTGGT 140

RESULT 3
CA795230
LOCUS
DEFINITION
Cac BL 226 Cac BL (Bean and Leaf from Amelonardo type Cacao)
EST 05-DEC-2002
Theobroma cacao cDNA clone Cac_BL_226 5', mRNA sequence.
ACCESSION
CA795230
VERSION
CA795230.1 GI:26052306
KEYWORDS
SOURCE
Theobroma cacao (cacao)
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;
Theobroma.
REFERENCE
1 (bases 1 to 338)
Jones,P.G., Allaway,D., Gilmour,D.M., Harris,C., Rankin,D.,
Retzel,E.R. and Jones,C.A.
Gene discovery and microarray analysis of cacao (Theobroma cacao
L.) varieties
Planta 216 (2), 255-264 (2002)
22337596
12447539
PUBMED
COMMENT
Contact: Jones, Paul
Masterfoods
3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
Tel: +44 1664 416644
Email: Paul.Jones@eu.affem.com
Seq primer: T3.
FEATURES
source
1..338
/organism="Theobroma cacao"
/mol_type="mRNA"
/strain="Amelonado type"
/db_xref="taxon:3641"
/clone="Cac BL 226"
/tissue_type="Mature leaf and mature bean"
/cell_type="Whole organ"
/dev_stage="maturity"
/lab_host="XL-1 Blue MRF"
/clone_lib="Cac_BL (Bean and Leaf from Amelonardo type
Cacao)"
/note="Vector: pBK-CMV; Bean and leaf tissue from an
Amelonado type Cacao tree."

```

```

ORIGIN
Alignment Scores:
Pred. No.: 4.26e-09 Length: 338
Score: 117.00 Matches: 23
Percent Similarity: 88.46% Conservative: 0
Best Local Similarity: 88.46% Mismatches: 3
Query Match: 95.90% Indels: 0
DB: 6 Gaps: 0

```

US-10-691-590-3 (1-26) x CA7955230 (1-338)

QY 1 ArgArgGluGlnGluGluSerGluGluThrPheGlyGluPhe\*\*\*GlnVal\*\*\* 20  
|||||  
Db 93 AGAAGAGAACAGAGAGAGTCAAGAGAGGACATTTGGAGAAATTCACAGAGTCAAA 152  
|||||  
QY 21 AlaProLeu\*\*\*ProGly 26  
|||||  
Db 153 GCCCCATTGTCACTGGT 170  
|||||

## RESULT 4

CA795596 351 bp mRNA linear EST 05-DEC-2002  
LOCUS  
DEFINITION  
Cac BL 2623 Cac BL (Bean and Leaf from Amelonardo type Cacao)

Theobroma cacao cDNA clone Cac\_BL\_2623 5', mRNA sequence.

ACCESSION

CA795596

VERSION

CA795596.1 GI:26052672

KEYWORDS

EST.

SOURCE

Theobroma cacao (cacao)

ORGANISM

Theobroma cacao

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;

Theobroma.

REFERENCE

1 (bases 1 to 351)

AUTHORS

Jones, P.G., Allaway, D., Gilmour, D.M., Harris, C., Rankin, D.,

Retzel, E.R. and Jones, C.A.

Gene discovery and microarray analysis of cacao (Theobroma cacao

L.) varieties

Planta 216 (2), 255-264 (2002)

PUBMED

12447539

COMMENT

Contact: Jones, Paul

Masterfoods

3d Dundee Road, Slough, Berkshire, UK, SL1 4LG

Tel: +44 1664 416644

Email: Paul.Jones@eu.affem.com

Seq primer: T3.

Location/Qualifiers

1. .351

/organism="Theobroma cacao"

/mol\_type="mRNA"

/strain="Amelonado type"

/db\_xref="taxon:3641"

/clone="Cac\_BL\_2623"

/tissue\_type="Mature leaf and mature bean"

/cell\_type="Whole organ"

/dev\_stage="maturity"

/lab\_host="XL-1 Blue MRF"

/clone\_lib="Cac\_BL (Bean and Leaf from Amelonardo type

Cacao)"

/note="Vector: pBK-CMV; Bean and leaf tissue from an

Amelonado type Cacao tree."

ORIGIN

Alignment Scores:

Pred. No.: 4.44e-09

Score: 117.00

Length: 351

Matches: 23

Conservative: 0

Percent Similarity: 88.46%

Best Local Similarity: 88.46%

Mismatch: 3

Query Match: 95.90%

Indels: 0

Gaps: 0

## RESULT 5

CA796844

LOCUS

DEFINITION

Cac BL 3891 Cac BL (Bean and Leaf from Amelonardo type Cacao)

Theobroma cacao cDNA clone Cac\_BL\_3891 5', mRNA sequence.

ACCESSION

CA796844

VERSION

CA796844.1 GI:26053930

KEYWORDS

EST.

SOURCE

Theobroma cacao (cacao)

ORGANISM

Theobroma cacao

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;

Theobroma.

REFERENCE

1 (bases 1 to 359)

AUTHORS

Jones, P.G., Allaway, D., Gilmour, D.M., Harris, C., Rankin, D.,

Retzel, E.R. and Jones, C.A.

Gene discovery and microarray analysis of cacao (Theobroma cacao

L.) varieties

Planta 216 (2), 255-264 (2002)

PUBMED

12447539

COMMENT

Contact: Jones, Paul

Masterfoods

3d Dundee Road, Slough, Berkshire, UK, SL1 4LG

Tel: +44 1664 416644

Email: Paul.Jones@eu.affem.com

Seq primer: T3.

Location/Qualifiers

1. .359

/organism="Theobroma cacao"

/mol\_type="mRNA"

/strain="Amelonado type"

/db\_xref="taxon:3641"

/clone="Cac\_BL\_3891"

/tissue\_type="Mature leaf and mature bean"

/cell\_type="Whole organ"

/dev\_stage="maturity"

/lab\_host="XL-1 Blue MRF"

/clone\_lib="Cac\_BL (Bean and Leaf from Amelonardo type

Cacao)"

/note="Vector: pBK-CMV; Bean and leaf tissue from an

Amelonado type Cacao tree."

ORIGIN

Alignment Scores:

Pred. No.: 4.56e-09

Score: 117.00

Length: 359

Matches: 23

Conservative: 0

Percent Similarity: 88.46%

Best Local Similarity: 88.46%

Mismatch: 3

Query Match: 95.90%

Indels: 0

Gaps: 0

US-10-691-590-3 (1-26) x CA796844 (1-359)

QY 1 ArgArgGluGlnGluGluSerGluGluThrPheGlyGluPhe\*\*\*GlnVal\*\*\* 20

|||||

Db 90 AGAAGAGAACAGAGAGAGTCAAGAGAGGACATTTGGAGAAATTCACAGAGTCAAA 149

|||||

QY 21 AlaProLeu\*\*\*ProGly 26

|||||

Db 150 GCCCCATTGTCACTGGT 167

|||||

## RESULT 6

CA796065

LOCUS

DEFINITION

Cac BL 3076 Cac BL (Bean and Leaf from Amelonardo type Cacao)

Theobroma cacao cDNA clone Cac\_BL\_3076 5', mRNA sequence.

ACCESSION

CA796065

VERSION

CA796065.1 GI:26053141

KEYWORDS

EST.

SOURCE

Theobroma cacao (cacao)

ORGANISM

Theobroma cacao

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;

Theobroma.

REFERENCE

1 (bases 1 to 362)

AUTHORS

Jones, P.G., Allaway, D., Gilmour, D.M., Harris, C., Rankin, D.,

Retzel, E.R. and Jones, C.A.

Gene discovery and microarray analysis of cacao (Theobroma cacao

L.) varieties

Planta 216 (2), 255-264 (2002)

PUBMED

12447539

COMMENT

Contact: Jones, Paul

Masterfoods

3d Dundee Road, Slough, Berkshire, UK, SL1 4LG

Tel: +44 1664 416644

Email: Paul.Jones@eu.affem.com

Seq primer: T3.

Location/Qualifiers

1. .359

/organism="Theobroma cacao"

/mol\_type="mRNA"

/strain="Amelonado type"

/db\_xref="taxon:3641"

/clone="Cac\_BL\_3891"

/tissue\_type="Mature leaf and mature bean"

/cell\_type="Whole organ"

/dev\_stage="maturity"

/lab\_host="XL-1 Blue MRF"

/clone\_lib="Cac\_BL (Bean and Leaf from Amelonardo type

Cacao)"

/note="Vector: pBK-CMV; Bean and leaf tissue from an

Amelonado type Cacao tree."

ORIGIN

Alignment Scores:

Pred. No.: 4.44e-09

Score: 117.00

Length: 351

Matches: 23

Conservative: 0

Percent Similarity: 88.46%

Best Local Similarity: 88.46%

Mismatch: 3

Query Match: 95.90%

Indels: 0

Gaps: 0

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae; Theobroma.

REFERENCE  
AUTHORS Jones, P.G., Allaway, D., Gilmour, D.M., Harris, C., Rankin, D., Retzel, E.R. and Jones, C.A.  
TITLE Gene discovery and microarray analysis of cacao (Theobroma cacao L.) varieties

JOURNAL Planta 216 (2), 255-264 (2002)  
MEDLINE 22337596  
PUBMED 12447539  
COMMENT Masterfoods  
3d Dundee Road, Slough, Berkshire, UK, SL1 4LG  
Tel: +44 1664 416644  
Email: Paul.Jones@eu.affem.com  
Seq primer: T3.

FEATURES  
source  
1..362  
/organism="Theobroma cacao"  
/mol\_type="mRNA"  
/strain="Amelonado type"  
/db\_xref="taxon:3641"  
/clone="Cac.BL.3076"  
/tissue\_type="Mature leaf and mature bean"  
/cell\_type="Whole organ"  
/dev\_stage="maturity"  
/lab\_host="XL-1 Blue MRF"  
/clone\_lib="Cac.BL (Bean and Leaf from Amelonado type Cacao)"  
/note="Vector: pBK-CMV; Bean and leaf tissue from an Amelonado type Cacao tree."

## ORIGIN

Alignment Scores:  
Pred. No.: 4 6e-09 Length: 362  
Score: 117.00 Matches: 23  
Percent Similarity: 88.46% Conservative: 0  
Best Local Similarity: 88.46% Mismatches: 3  
Query Match: 95.90% Indels: 0  
DB: 6 Gaps: 0

US-10-691-590-3 (1-26) x CA796065 (1-362)

QY 1 ArgArgGluGlnGluSerGluGluThrPheGlyGluPhe\*\*\*GlnVal\*\*\* 20  
|||||  
Db 171 AGAAGAGACAAAGAGAGAGTCAAGAGAGGAGACATTTGGAGAAATCCAGCAGGTCAAA 230  
|||||

QY 21 AlaProLeu\*\*\*ProGly 26  
|||||  
Db 231 GCCCATTTGTCACCTGGT 248  
|||||

RESULT 7  
CA797330 364 bp mRNA linear EST 05-DEC-2002  
LOCUS Cac.BL.442 Cac.BL (Bean and Leaf from Amelonado type Cacao)  
DEFINITION Theobroma cacao cDNA clone Cac.BL.442 5', mRNA sequence.

ACCESSION CA797330  
VERSION CA797330.1 GI:26054416  
KEYWORDS EST.  
SOURCE Theobroma cacao (cacao)  
ORGANISM Theobroma cacao  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae; Theobroma.

REFERENCE  
AUTHORS Jones, P.G., Allaway, D., Gilmour, D.M., Harris, C., Rankin, D., Retzel, E.R. and Jones, C.A.  
TITLE Gene discovery and microarray analysis of cacao (Theobroma cacao L.) varieties  
JOURNAL Planta 216 (2), 255-264 (2002)

MEDLINE 22337596  
PUBMED 12447539  
COMMENT Contact: Jones, Paul  
Masterfoods  
3d Dundee Road, Slough, Berkshire, UK, SL1 4LG  
Tel: +44 1664 416644  
Email: Paul.Jones@eu.affem.com  
Seq primer: T3.

FEATURES  
source  
1..364  
/organism="Theobroma cacao"  
/mol\_type="mRNA"  
/strain="Amelonado type"  
/db\_xref="taxon:3641"  
/clone="Cac.BL.442"  
/tissue\_type="Mature leaf and mature bean"  
/cell\_type="Whole organ"  
/dev\_stage="maturity"  
/lab\_host="XL-1 Blue MRF"  
/clone\_lib="Cac.BL (Bean and Leaf from Amelonado type Cacao)"  
/note="Vector: pBK-CMV; Bean and leaf tissue from an Amelonado type Cacao tree."

## ORIGIN

Alignment Scores:  
Pred. No.: 4 63e-09 Length: 364  
Score: 117.00 Matches: 23  
Percent Similarity: 88.46% Conservative: 0  
Best Local Similarity: 88.46% Mismatches: 3  
Query Match: 95.90% Indels: 0  
DB: 6 Gaps: 0

US-10-691-590-3 (1-26) x CA797330 (1-364)

QY 1 ArgArgGluGlnGluSerGluGluThrPheGlyGluPhe\*\*\*GlnVal\*\*\* 20  
|||||  
Db 186 AGAAGAGACAAAGAGAGAGTCAAGAGAGGAGACATTTGGAGAAATCCAGCAGGTCAAA 245  
|||||  
QY 21 AlaProLeu\*\*\*ProGly 26  
|||||  
Db 246 GCCCATTTGTCACCTGGT 263  
|||||

## RESULT 8

CA795334 378 bp mRNA linear EST 05-DEC-2002  
LOCUS Cac.BL.2374 Cac.BL (Bean and Leaf from Amelonado type Cacao)  
DEFINITION Theobroma cacao cDNA clone Cac.BL.2374 5', mRNA sequence.

ACCESSION CA795334  
VERSION CA795334.1 GI:26052410  
KEYWORDS EST.  
SOURCE Theobroma cacao (cacao)  
ORGANISM Theobroma cacao  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae; Theobroma.

REFERENCE  
AUTHORS Jones, P.G., Allaway, D., Gilmour, D.M., Harris, C., Rankin, D., Retzel, E.R. and Jones, C.A.

TITLE Gene discovery and microarray analysis of cacao (Theobroma cacao L.) varieties  
JOURNAL Planta 216 (2), 255-264 (2002)  
MEDLINE 22337596  
PUBMED 12447539  
COMMENT Contact: Jones, Paul  
Masterfoods  
3d Dundee Road, Slough, Berkshire, UK, SL1 4LG  
Tel: +44 1664 416644  
Email: Paul.Jones@eu.affem.com  
Seq primer: T3.

FEATURES  
source  
1..378  
Location/Qualifiers

```

/organism="Theobroma cacao"
/mol_type="mRNA"
/strain="Amelonado type"
/db_xref="taxon:3641"
/clone="Cac BL 2374"
/tissue_type="Mature leaf and mature bean"
/cell_type="Whole organ"
/dev_stage="maturity"
/lab_host="XL-1 Blue MRF"
/clone_lib="Cac_BL (Bean and Leaf from Amelonado type Cacao)"
/notes="Vector: pBK-CMV; Bean and leaf tissue from an Amelonado type Cacao tree."

```

ORIGIN

Alignment Scores:

Pred. No.:	4.83e-09	Length:	378
Score:	117.00	Matches:	23
Percent Similarity:	88.46%	Conservative:	0
Best Local Similarity:	88.46%	Mismatches:	3
Query Match:	95.90%	Indels:	0
DB:	6	Gaps:	0

US-10-691-590-3 (1-26) x CA795334 (1-378)

```

QY 1 ArgArgGluGlnGluSerGluGluThrPheGlyGluPhe***GinVal*** 20
    |||||
Db 243 AGAAGAGAACAGAGAGAGTCAAGAGAGGAGACATTGGAGAAATCCAGCAGGTCAAA 302
    |||||

QY 21 AlaProLeu***ProGly 26
    |||||
Db 303 GCCCATTGTCACCTGGT 320
    |||||

```

RESULT 9

CA796843

LOCUS

DEFINITION

CA796843

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Theobroma cacao (cacao)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae; Theobroma.

1 (bases 1 to 423)

Jones, P.G., Allaway, D., Gilmour, D.M., Harris, C., Rankin, D., Retzel, E.R. and Jones, C.A.

Gene discovery and microarray analysis of cacao (Theobroma cacao L.) varieties

Planta 216 (2), 255-264 (2002)

22337596

12447539

Contact: Jones, Paul

Masterfoods

3d Dundee Road, Slough, Berkshire, UK, SL1 4LG

Tel: +44 1664 416644

Email: Paul.Jones@eu.affem.com

Seq primer: T3.

FEATURES

source

```

1. .423
/organism="Theobroma cacao"
/mol_type="mRNA"
/strain="Amelonado type"
/db_xref="taxon:3641"
/clone="Cac BL 3890"
/tissue_type="Mature leaf and mature bean"
/cell_type="Whole organ"
/dev_stage="maturity"
/lab_host="XL-1 Blue MRF"
/clone_lib="Cac_BL (Bean and Leaf from Amelonado type Cacao)"

```

Alignment Scores:

Pred. No.:	5.57e-09	Length:	430
Score:	117.00	Matches:	23
Percent Similarity:	88.46%	Conservative:	0
Best Local Similarity:	88.46%	Mismatches:	3

```

Cacao)"
/notes="Vector: pBK-CMV; Bean and leaf tissue from an Amelonado type Cacao tree."

```

ORIGIN

Alignment Scores:

Pred. No.:	5.46e-09	Length:	423
Score:	117.00	Matches:	23
Percent Similarity:	88.46%	Conservative:	0
Best Local Similarity:	88.46%	Mismatches:	3
Query Match:	95.90%	Indels:	0
DB:	6	Gaps:	0

US-10-691-590-3 (1-26) x CA796843 (1-423)

```

QY 1 ArgArgGluGlnGluSerGluGluThrPheGlyGluPhe***GinVal*** 20
    |||||
Db 269 AGAAGAGAACAGAGAGAGTCAAGAGAGGAGACATTGGAGAAATCCAGCAGGTCAAA 328
    |||||

QY 21 AlaProLeu***ProGly 26
    |||||
Db 329 GCCCATTGTCACCTGGT 346
    |||||

```

RESULT 10

CA794888

LOCUS

DEFINITION

CA794888

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Theobroma cacao (cacao)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae; Theobroma.

1 (bases 1 to 430)

Jones, P.G., Allaway, D., Gilmour, D.M., Harris, C., Rankin, D., Retzel, E.R. and Jones, C.A.

Gene discovery and microarray analysis of cacao (Theobroma cacao L.) varieties

Planta 216 (2), 255-264 (2002)

22337596

12447539

Contact: Jones, Paul

Masterfoods

3d Dundee Road, Slough, Berkshire, UK, SL1 4LG

Tel: +44 1664 416644

Email: Paul.Jones@eu.affem.com

Seq primer: T3.

FEATURES

source

```

1. .430
/organism="Theobroma cacao"
/mol_type="mRNA"
/strain="Amelonado type"
/db_xref="taxon:3641"
/clone="Cac BL 1941"
/tissue_type="Mature leaf and mature bean"
/cell_type="Whole organ"
/dev_stage="maturity"
/lab_host="XL-1 Blue MRF"
/clone_lib="Cac_BL (Bean and Leaf from Amelonado type Cacao)"
/notes="Vector: pBK-CMV; Bean and leaf tissue from an Amelonado type Cacao tree."

```

ORIGIN

Alignment Scores:

Pred. No.:	5.57e-09	Length:	430
Score:	117.00	Matches:	23
Percent Similarity:	88.46%	Conservative:	0
Best Local Similarity:	88.46%	Mismatches:	3

```

Query Match: 95.90% Indels: 0
DB: 6 Gaps: 0

US-10-691-590-3 (1-26) x CA794888 (1-430)

QY 1 ArgArgGluGlnGluGluSerGluGluThrPheGlyGluPhe***GlnVal*** 20
|||||
Db 271 AGAAGAGAACAGAGAGAGTCAAGAGGAGACATTTGGAGAAATTCACAGCAGGTCAAA 330
|||||

QY 21 AlaProLeu***ProGly 26
|||||
Db 331 GCCCATTGTCACCTGGT 348
|||||

RESULT 11
CA797867
LOCUS
DEFINITION
CA797867 441 bp mRNA linear EST 05-DEC-2002
Theobroma cacao cDNA clone Cac_BL_51 5', mRNA sequence.
ACCESSION
CA797867
KEYWORDS
EST.
SOURCE
Theobroma cacao (cacao)
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;
Theobroma.
REFERENCE
1 (bases 1 to 441)
JONES, P.G., ALLAWAY, D., GILMOUR, D.M., HARRIS, C., RANKIN, D.,
RETZEL, E.R. and JONES, C.A.
Gene discovery and microarray analysis of cacao (Theobroma cacao
L.) varieties
Planta 216 (2), 255-264 (2002)
PUBMED
12447539
COMMENT
Contact: Jones, Paul
Masterfoods
3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
Tel: +44 1664 416644
Email: Paul.Jones@eu.affem.com
Seq primer: T3.
FEATURES
source
Location/Qualifiers
1..441
/organism="Theobroma cacao"
/mol_type="mRNA"
/strain="Amelonado type"
/db_xref="taxon:3641"
/clone="Cac BL 51"
/tissue_type="Mature leaf and mature bean"
/cell_type="whole organ"
/dev_stage="maturity"
/lab_host="XL-1 Blue MRF"
/clone_lib="Cac BL (Bean and Leaf from Amelonado type
Cacao)"
/note="Vector: pBK-CMV; Bean and leaf tissue from an
Amelonado type Cacao tree."
ORIGIN
Alignment Scores:
Pred. No.: 5.72e-09 Length: 441
Score: 117.00 Matches: 23
Percent Similarity: 88.46% Conservative: 0
Best Local Similarity: 88.46% Mismatches: 3
Query Match: 95.90% Indels: 0
DB: 6 Gaps: 0

US-10-691-590-3 (1-26) x CA797867 (1-441)

QY 1 ArgArgGluGlnGluGluSerGluGluThrPheGlyGluPhe***GlnVal*** 20
|||||
Db 318 AGAAGAGAACAGAGAGAGTCAAGAGGAGACATTTGGAGAAATTCACAGCAGGTCAAA 377
|||||

QY 21 AlaProLeu***ProGly 26
|||||

US-10-691-590-3 (1-26) x CA795681 (1-463)

QY 1 ArgArgGluGlnGluGluSerGluGluThrPheGlyGluPhe***GlnVal*** 20
|||||
Db 282 AGAAGAGAACAGAGAGAGTCAAGAGGAGACATTTGGAGAAATTCACAGCAGGTCAAA 341
|||||

QY 21 AlaProLeu***ProGly 26
|||||
Db 342 GCCCATTGTCACCTGGT 359
|||||

RESULT 13
CA794398
LOCUS
DEFINITION
CA794398 465 bp mRNA linear EST 05-DEC-2002
Theobroma cacao cDNA clone Cac_BL_1631 5', mRNA sequence.
ACCESSION
CA794398
VERSION
CA794398.1 GI:26051474

```



```
KEYWORDS EST.
SOURCE Theobroma cacao (cacao)
ORGANISM Theobroma cacao

REFERENCE 1 (bases 1 to 465)
AUTHORS Jones P.G., Allaway, D., Gilmour, D.M., Harris, C., Rankin, D.,
Retzel, E.R. and Jones, C.A.
TITLE Gene discovery and microarray analysis of cacao (Theobroma cacao
L.) varieties
JOURNAL Planta 216 (2), 255-264 (2002)
MEDLINE 22337596
PUBMED 12447539
COMMENT Contact: Jones, Paul
Masterfoods
3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
Tel: +44 1664 416644
Email: Paul.Jones@eu.affem.com
Seq primer: T3
Location/Qualifiers
source 1. .465
/organism="Theobroma cacao"
/mol_type="mRNA"
/strain="Amelonado type"
/db_xref="taxon:3641"
/clone="Cac_BL_1631"
/tissue_type="Mature leaf and mature bean"
/cell_type="Whole organ"
/dev_stage="maturity"
/lab_host="XL-1 Blue MRF"
/clone_lib="Cac_BL (Bean and Leaf from Amelonado type
Cacao)"
/note="Vector: pBK-CMV; Bean and leaf tissue from an
Amelonado type Cacao tree."

ORIGIN
Alignment Scores: 6.07e-09 Length: 465
Pred. No.: 117.00 Matches: 23
Score: 88.46% Conservative: 0
Percent Similarity: 88.46% Mismatches: 3
Best Local Similarity: 88.46% Indels: 0
Query Match: 95.90% Gaps: 0
DB: 6

US-10-691-590-3 (1-26) x CA794398 (1-465)

QY 1 ArgArgGluGlnGluGluSerGluGluThrPheGlyGluPhe***GlnVal*** 20
|||||
Db 358 AGAAGAGAACAAAGAGAGAGTCAGAGAGAGACATTTGGAGAAATTCACAGCAGGTCAA 417

QY 21 AlaProLeu***ProGly 26
|||||
Db 418 GCCCATGTGTCACCTGCT 435

RESULT 14
CA797010
LOCUS CA797010 468 bp mRNA linear EST 05-DEC-2002
DEFINITION Cac BL 4090 Cac BL (Bean and Leaf from Amelonado type Cacao)
Theobroma cacao cDNA clone Cac_BL_4090 5', mRNA sequence.
ACCESSION CA797010
VERSION CA797010.1 GI:26054096
KEYWORDS EST.
SOURCE Theobroma cacao (cacao)
ORGANISM Theobroma cacao
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;
Theobroma.
1 (bases 1 to 468)
Jones, P.G., Allaway, D., Gilmour, D.M., Harris, C., Rankin, D.,
Retzel, E.R. and Jones, C.A.

KEYWORDS EST.
SOURCE Theobroma cacao (cacao)
ORGANISM Theobroma cacao
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;
Theobroma.
1 (bases 1 to 468)
Jones, P.G., Allaway, D., Gilmour, D.M., Harris, C., Rankin, D.,
Retzel, E.R. and Jones, C.A.
```

```
TITLE Gene discovery and microarray analysis of cacao (Theobroma cacao
L.) varieties
JOURNAL Planta 216 (2), 255-264 (2002)
MEDLINE 22337596
PUBMED 12447539
COMMENT Contact: Jones, Paul
Masterfoods
3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
Tel: +44 1664 416644
Email: Paul.Jones@eu.affem.com
Seq primer: T3
Location/Qualifiers
source 1. .468
/organism="Theobroma cacao"
/mol_type="mRNA"
/strain="Amelonado type"
/db_xref="taxon:3641"
/clone="Cac_BL_4090"
/tissue_type="Mature leaf and mature bean"
/cell_type="Whole organ"
/dev_stage="maturity"
/lab_host="XL-1 Blue MRF"
/clone_lib="Cac_BL (Bean and Leaf from Amelonado type
Cacao)"
/note="Vector: pBK-CMV; Bean and leaf tissue from an
Amelonado type Cacao tree."

ORIGIN
Alignment Scores: 6.11e-09 Length: 468
Pred. No.: 117.00 Matches: 23
Score: 88.46% Conservative: 0
Percent Similarity: 88.46% Mismatches: 3
Best Local Similarity: 88.46% Indels: 0
Query Match: 95.90% Gaps: 0
DB: 6

US-10-691-590-3 (1-26) x CA797010 (1-468)

QY 1 ArgArgGluGlnGluGluSerGluGluThrPheGlyGluPhe***GlnVal*** 20
|||||
Db 352 AGAAGAGAACAAAGAGAGAGTCAGAGAGAGACATTTGGAGAAATTCACAGCAGGTCAA 411

QY 21 AlaProLeu***ProGly 26
|||||
Db 412 GCCCATGTGTCACCTGCT 429

RESULT 15
CA795491
LOCUS CA795491 474 bp mRNA linear EST 05-DEC-2002
DEFINITION Cac BL_2524 Cac BL (Bean and Leaf from Amelonado type Cacao)
Theobroma cacao cDNA clone Cac_BL_2524 5', mRNA sequence.
ACCESSION CA795491
VERSION CA795491.1 GI:26052567
KEYWORDS EST.
SOURCE Theobroma cacao (cacao)
ORGANISM Theobroma cacao
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;
Theobroma.
1 (bases 1 to 474)
Jones, P.G., Allaway, D., Gilmour, D.M., Harris, C., Rankin, D.,
Retzel, E.R. and Jones, C.A.
Gene discovery and microarray analysis of cacao (Theobroma cacao
L.) varieties
JOURNAL Planta 216 (2), 255-264 (2002)
MEDLINE 22337596
PUBMED 12447539
COMMENT Contact: Jones, Paul
Masterfoods
3d Dundee Road, Slough, Berkshire, UK, SL1 4LG
Tel: +44 1664 416644
Email: Paul.Jones@eu.affem.com
```

```
FEATURES
Source
Seq primer: T3.
Location/Qualifiers
1..474
/organism="Theobroma cacao"
/mol_type="mRNA"
/strain="Amelonado type"
/db_xref="taxon:3641"
/clone="Cac BL 2524"
/tissue_type="Mature leaf and mature bean"
/cell_type="Whole organ"
/dev_stage="maturity"
/lab_host="Xl-1 Blue MRF"
/clone_lib="Cac BL (Bean and Leaf from Amelonado type
Cacao)"
/note="Vector: pBK-CMV; Bean and leaf tissue from an
Amelonado type Cacao tree."

ORIGIN
Alignment Scores:
Pred. No.: 6.2e-09 Length: 474
Score: 117.00 Matches: 23
Percent Similarity: 88.46% Conservative: 0
Best Local Similarity: 88.46% Mismatches: 3
Query Match: 95.90% Indels: 0
DB: 6 Gaps: 0

US-10-691-590-3 (1-26) x CA795491 (1-474)
QY 1 ArgArgGluGluGluSerGluGluThrPheGlyGluPhe**GlnVal*** 20
DB 247 AGAAGAGAACAGAGAGAGTCAAGAGAGGAGACATTTGGAGATTCCAGCAGGTCAAA 306
QY 21 AlaProLeu***ProGly 26
DB 307 GCCCATTTGTCACCTGGT 324

Search completed: November 5, 2005, 05:56:57
Job time : 1032.47 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 5, 2005, 02:54:27 ; Search time 263.165 Seconds  
(without alignments)  
817.035 Million cell updates/sec

Title: US-10-691-590-3  
Perfect score: 122  
Sequence: 1 RREQEESEETFGFQVXAPLXPG 26

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 9794790 seqs, 4134909567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US10691590/runat\_02112005\_120023\_13233/app\_query.fasta\_1.462  
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS-human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US10691590@cgn\_1\_582@runat\_02112005\_120023\_13233  
-NCPU=6 -ICPU=3 -NO MAP -LARGUEURY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=130 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq.\*  
18: /cgn2\_6/ptodata/2/pubpna/US10E\_PUBCOMB.seq.\*  
19: /cgn2\_6/ptodata/2/pubpna/US10F\_PUBCOMB.seq.\*  
20: /cgn2\_6/ptodata/2/pubpna/US10G\_PUBCOMB.seq.\*  
21: /cgn2\_6/ptodata/2/pubpna/US10H\_PUBCOMB.seq.\*  
22: /cgn2\_6/ptodata/2/pubpna/US10I\_PUBCOMB.seq.\*  
23: /cgn2\_6/ptodata/2/pubpna/US10J\_PUBCOMB.seq.\*  
24: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
25: /cgn2\_6/ptodata/2/pubpna/US11A\_PUBCOMB.seq.\*  
26: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*  
27: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
28: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	61	50.0	2362	20	US-10-437-963-58448
C 2	59	48.4	1845	17	US-10-029-386-24791
C 3	59	48.4	2365	21	US-10-723-860-1651
C 4	59	48.4	2365	22	US-10-741-600-24
C 5	59	48.4	2375	22	US-10-741-600-23
C 6	59	48.4	2382	21	US-10-723-860-6048
C 7	59	48.4	14123	22	US-10-741-600-17921
C 8	59	48.4	16220	22	US-10-741-600-17562
C 9	59	48.4	43623	22	US-10-741-600-17818
C 10	57	46.7	584	21	US-10-425-115-160535
C 11	57	46.7	1047	20	US-10-437-963-60949
C 12	55	45.1	2589	26	US-11-097-143-11710
C 13	54.5	44.7	4754	9	US-09-982-091A-1
C 14	54	44.3	575	13	US-09-925-065A-133584
C 15	54	44.3	625	13	US-09-925-065A-854915
C 16	54	44.3	658	13	US-09-925-065A-93913
C 17	54	44.3	684	20	US-10-437-963-2415
C 18	54	44.3	752	19	US-10-424-599-31267
C 19	54	44.3	11817	16	US-10-156-761-2884
C 20	54	44.3	65253	18	US-10-085-117-331
C 21	54	44.3	125746	16	US-10-156-761-15102
C 22	54	44.3	9025608	16	US-10-156-761-1
C 23	53	43.4	316	20	US-10-674-124A-12347
C 24	53	43.4	727	21	US-10-425-115-90035
C 25	53	43.4	1017	20	US-10-437-963-45963
C 26	53	43.4	1301	21	US-10-739-930-1674
C 27	53	43.4	4756	9	US-09-982-091A-3
C 28	53	43.4	4804	21	US-10-357-930-30217
C 29	53	43.4	111206	14	US-10-087-192-1957
C 30	53	43.4	134841	14	US-10-087-192-1987
C 31	52.5	43.0	1281	20	US-10-437-963-15584
C 32	52	42.6	403	19	US-10-276-774-259
C 33	52	42.6	405	9	US-09-770-423-713
C 34	52	42.6	698	14	US-10-027-632-148019
C 35	52	42.6	698	18	US-10-027-632-148019
C 36	52	42.6	1297	19	US-10-425-114-29320
C 37	52	42.6	1540	20	US-10-437-963-5733
C 38	52	42.6	1580	19	US-10-425-114-20659
C 39	52	42.6	1619	19	US-10-424-599-75133
C 40	52	42.6	1874	19	US-10-424-599-75130
C 41	52	42.6	2667	19	US-10-424-599-91404
C 42	52	42.6	147309	9	US-09-742-312-3
C 43	52	42.6	147309	17	US-10-436-185-3
C 44	52	42.6	312477	20	US-10-317-883A-12
C 45	51.5	42.2	6953	19	US-10-600-230-66

# ALIGNMENTS

## RESULT 1

US-10-437-963-58448/c  
; Sequence 58448, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovacic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With



.; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF



```

; ORGANISM: Xenopus laevis
US-09-982-091A-1

Alignment Scores:
Pred. No.: 38.5 Length: 4754
Score: 54.50 Matches: 12
Percent Similarity: 86.67% Conservative: 1
Best Local Similarity: 80.00% Mismatches: 1
Query Match: 44.67% Indels: 1
DB: 9 Gaps: 1

US-10-691-590-3 (1-26) x US-09-982-091A-1 (1-4754)

Qy 2 ArgGluGluGluSerGluGluGluThrPheGlyGluPhe 16
|||:|||||:|||||
Db 2885 AGAGAAACAGGAAGATGAGGAAGAAGAA---TTGGTGAATTC 2926

RESULT 14
US-09-925-065A-133584
; Sequence 133584, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 133584
; LENGTH: 575
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-133584

Alignment Scores:
Pred. No.: 4.79 Length: 575
Score: 54.00 Matches: 10
Percent Similarity: 91.67% Conservative: 1
Best Local Similarity: 83.33% Mismatches: 1
Query Match: 44.26% Indels: 0
DB: 13 Gaps: 0

US-10-691-590-3 (1-26) x US-09-925-065A-133584 (1-575)

Qy 3 GluGluGluGluSerGluGluGluThrPheGly 14
|||:|||||:|||||
Db 88 GAGGAGACAGAGAGAGAGGAGGAGAGACATTGGT 123

RESULT 15
US-09-925-065A-854915
; Sequence 854915, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24

```





GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 5, 2005, 02:34:02 ; Search time 56.2957 Seconds  
(without alignments)

755.710 Million cell updates/sec

Title: US-10-691-590-3

Perfect score: 122

Sequence: 1 RREQEESEETFGFXQVXAPLXPG 26

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US10691590/runat\_02112005\_120023\_13203/app\_query.fasta\_1.462  
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10691590 @cgn\_1\_93 @runat\_02112005\_120023\_13203 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCKS=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA.\*

1: /cgn2\_6/ptodata/1/ina/5A COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PTUS COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	117	95.9	1867	1	US-07-955-905A-1
2	53	43.4	4756	4	US-09-949-016-4455
3	51	41.8	53737	4	US-09-949-016-16197
4	50.5	41.4	601	4	US-09-949-016-92026
5	50.5	41.4	254405	4	US-09-949-016-14381
6	50	41.0	1816	4	US-09-762-258-1
7	49	40.2	601	4	US-09-949-016-65666
8	49	40.2	2677	4	US-09-949-016-1924
9	49	40.2	4474	3	US-09-221-017B-114
10	49	40.2	19237	4	US-09-949-016-13666
11	49	40.2	19237	4	US-09-949-016-13666
12	48	39.3	330	4	US-09-489-039A-4753

c 13	48	39.3	420	4	US-09-489-039A-4835	Sequence 4835, Ap
c 14	48	39.3	790	3	US-08-998-416-421	Sequence 421, App
c 15	48	39.3	1548	4	US-09-543-681A-3556	Sequence 3556, Ap
c 16	48	39.3	1741	3	US-08-795-430-12	Sequence 12, Appl
c 17	48	39.3	1741	3	US-09-355-700-12	Sequence 12, Appl
c 18	48	39.3	1741	4	US-09-534-376A-12	Sequence 12, Appl
c 19	48	39.3	9051	4	US-09-949-016-12112	Sequence 12112, A
c 20	48	39.3	9052	4	US-09-949-016-15529	Sequence 15529, A
c 21	48	39.3	94133	4	US-09-949-016-11901	Sequence 11901, A
c 22	48	39.3	94133	4	US-09-949-016-12713	Sequence 12713, A
c 23	48	39.3	94135	4	US-09-949-016-15934	Sequence 15934, A
c 24	48	39.3	94135	4	US-09-949-016-15935	Sequence 15935, A
c 25	48	39.3	94135	4	US-09-949-016-15936	Sequence 15936, A
c 26	48	39.3	94135	4	US-09-949-016-15937	Sequence 15937, A
c 27	48	39.3	109250	4	US-09-949-016-12530	Sequence 12530, A
c 28	48	39.3	109251	4	US-09-949-016-17321	Sequence 17321, A
c 29	48	39.3	237510	4	US-09-949-016-14273	Sequence 14273, A
c 30	47.5	38.9	11558	4	US-09-949-016-14317	Sequence 14317, A
c 31	47	38.5	56	4	US-09-981-803-24	Sequence 24, Appl
c 32	47	38.5	57	3	US-08-894-511-18	Sequence 18, Appl
c 33	47	38.5	57	3	US-09-655-728-18	Sequence 18, Appl
c 34	47	38.5	264	4	US-09-248-796A-6960	Sequence 6960, Ap
c 35	47	38.5	309	4	US-09-513-999C-20344	Sequence 20344, A
c 36	47	38.5	601	4	US-09-949-016-61851	Sequence 61851, A
c 37	47	38.5	601	4	US-09-949-016-112656	Sequence 112656, A
c 38	47	38.5	601	4	US-09-949-016-115267	Sequence 115267, A
c 39	47	38.5	601	4	US-09-949-016-144537	Sequence 144537, A
c 40	47	38.5	1053	4	US-09-248-796A-4335	Sequence 4335, Ap
c 41	47	38.5	1449	4	US-09-489-039A-2198	Sequence 2198, Ap
c 42	47	38.5	3937	4	US-10-164-595-7	Sequence 7, Appl
c 43	47	38.5	8355	3	US-08-406-030A-23	Sequence 23, Appl
c 44	47	38.5	9382	4	US-09-949-016-15306	Sequence 15306, A
c 45	47	38.5	44166	4	US-09-949-016-15829	Sequence 15829, A

#### ALIGNMENTS

RESULT 1  
US-07-955-905A-1  
; Sequence 1, Application US/07955905A  
; Patent No. 5770433  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: RECOMBINANT 47 AND 31 kD COCOA PROTEINS AND  
; PRECURSOR  
; NUMBER OF SEQUENCES: 28  
; COMPUTER READABLE FORM: disk  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/955,905A  
; FILING DATE: 21-JAN-1993  
; CLASSIFICATION: 435  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1867 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; ORIGINAL SOURCE: Theobroma cacao  
; ORGANISM: Theobroma cacao  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 14..1711  
; US-07-955-905A-1

Alignment Scores:  
Pred. No.: 1.4e-10  
Score: 117.00  
Percent Similarity: 88.46%  
Length: 1867  
Matches: 23  
Conservative: 0



```
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14381
; LENGTH: 254405
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14381

Alignment Scores:
Pred. No.: 9.33e+03 Length: 254405
Score: 50.50 Matches: 11
Percent Similarity: 53.12% Conservative: 6
Best Local Similarity: 34.38% Mismatches: 6
Query Match: 41.39% Indels: 9
DB: 4 Gaps: 1

US-10-691-590-3 (1-26) x US-09-949-016-14381 (1-254405)
QY 1 ArgArgGluGlnGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal***AlaProLeu 23
Db 93186 CAGAGAGAGAGACAGACGGGAAATGATATGACAGAGAGATGAATGGAAAACTGAAA 93127

QY 12 ThrPheGlyGluPhe***GlnVal***AlaProLeu 23
Db 93126 ACTTTGGAAATTTCCACCAAAATTAATACCCAGTT 93091

RESULT 6
US-09-762-258-1
; Sequence 1, Application US/09762258
; Patent No. 6830909
; GENERAL INFORMATION:
; APPLICANT: Gout, Ivan
; APPLICANT: Hara, Kenta
; APPLICANT: Waterfield, Michael
; APPLICANT: Yonezawa, Kazu
; TITLE OF INVENTION: Identification and Functional Characterization of a
; TITLE OF INVENTION: No. 6830909el Ribosomal S6 Protein Kinase
; FILE REFERENCE: 40750-5002-US
; CURRENT APPLICATION NUMBER: US/09/762,258
; CURRENT FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: PCT/US99/17595
; PRIOR FILING DATE: 1999-08-04
; PRIOR APPLICATION NUMBER: 60/095,268
; PRIOR FILING DATE: 1998-08-04
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1816
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (77)..(1561)
; OTHER INFORMATION: p70(beta) S6 Kinase gene
US-09-762-258-1

Alignment Scores:
Pred. No.: 37.3 Length: 1816
Score: 50.00 Matches: 10
Percent Similarity: 54.17% Conservative: 3
Best Local Similarity: 41.67% Mismatches: 11
Query Match: 40.98% Indels: 0
DB: 4 Gaps: 0

US-10-691-590-3 (1-26) x US-09-762-258-1 (1-1816)
```

```
QY 2 ArgGluGlnGluGluSerGluGluGluThrPheGlyGluPhe***GlnVal***Ala 21
Db 36 AGAGAGAGAGAGAGAGAGAGAGAGAGACTCGTCCGGAATGGCAGCGCGCGCGG 95

QY 22 ProLeu***Pro 25
Db 96 CCGCGGGGGCGC 107

RESULT 7
US-09-949-016-65666
; Sequence 65666, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 65666
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-65666

Alignment Scores:
Pred. No.: 15.4 Length: 601
Score: 49.00 Matches: 10
Percent Similarity: 80.00% Conservative: 2
Best Local Similarity: 66.67% Mismatches: 3
Query Match: 40.16% Indels: 0
DB: 4 Gaps: 0

US-10-691-590-3 (1-26) x US-09-949-016-65666 (1-601)
QY 1 ArgArgGluGlnGluGluSerGluGluGluThrPheGlyGlu 15
Db 498 CGAGGGGAGAGAGAGAGAGAGCGCGGAGAGAGCGGAGCGGAGGAG 542

RESULT 8
US-09-949-016-1924
; Sequence 1924, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1924
; LENGTH: 2677
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-1924

Alignment Scores:
```

Pred. No.: 86.6 Length: 2677  
Score: 49.00 Matches: 10  
Percent Similarity: 80.00% Conservatives: 2  
Best Local Similarity: 66.67% Mismatches: 3  
Query Match: 40.16% Indels: 0  
DB: 4 Gaps: 0

US-10-691-590-3 (1-26) x US-09-949-016-1924 (1-2677)

QY 1 ArgArgGluGlnGluGluSerGluGluGluThrPheGlyGlu 15  
|||:|||||:|||||:|||||:|||||:|||||  
Db 199 CGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 243

RESULT 9  
US-09-221-017B-114/c  
; Sequence 114, Application US/09221017B  
; Patent No. 6444799  
; GENERAL INFORMATION:  
; APPLICANT: ROSS, BRUCE C.  
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF  
; NUMBER OF SEQUENCES: 1120  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FORSTER  
; STREET: 755 PAGE MILL ROAD  
; CITY: PALO ALTO  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304-1018  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows  
; SOFTWARE: FastSeq for Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/221,017B  
; FILING DATE: 23-DEC-1998  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PP1182  
; FILING DATE: 31-DEC-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PP1546  
; FILING DATE: 30-JAN-1998  
; PRIOR APPLICATION DATA: PP2911  
; APPLICATION NUMBER: PP2911  
; FILING DATE: 09-APR-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/AU98/01023  
; FILING DATE: 10-DEC-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MONROY, GLADYS H  
; REGISTRATION NUMBER: 32,430  
; REFERENCE/DOCKET NUMBER: 27340-20021.00  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-813-5600  
; TELEFAX: 650-494-0792  
; TELEX: 706141  
; INFORMATION FOR SEQ ID NO: 114:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4474 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: circular  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: UNKNOWN  
; ORIGINAL SOURCE:  
; ORGANISM: PORPHYROMONAS GINGIVALIS  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 1...4474  
US-09-221-017B-114

Alignment Scores:  
Pred. No.: 157 Length: 4474  
Score: 49.00 Matches: 9  
Percent Similarity: 78.57% Conservatives: 2  
Best Local Similarity: 64.29% Mismatches: 3  
Query Match: 40.16% Indels: 0  
DB: 3 Gaps: 0

US-10-691-590-3 (1-26) x US-09-221-017B-114 (1-4474)

QY 1 ArgArgGluGlnGluGluSerGluGluGluThrPheGly 14  
|||:|||||:|||||:|||||:|||||:|||||  
Db 579 AGAGAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 538

RESULT 10  
US-09-949-016-13666  
; Sequence 13666, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. CRAIG ET AL.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13666  
; LENGTH: 19237  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-13666

Alignment Scores:  
Pred. No.: 848 Length: 19237  
Score: 49.00 Matches: 10  
Percent Similarity: 80.00% Conservatives: 2  
Best Local Similarity: 66.67% Mismatches: 3  
Query Match: 40.16% Indels: 0  
DB: 4 Gaps: 0

US-10-691-590-3 (1-26) x US-09-949-016-13666 (1-19237)

QY 1 ArgArgGluGlnGluGluSerGluGluGluThrPheGlyGlu 15  
|||:|||||:|||||:|||||:|||||:|||||  
Db 2199 CGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2243

RESULT 11  
US-09-949-016-13666/c  
; Sequence 13666, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. CRAIG ET AL.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13666

; LENGTH: 19237  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-13666

## Alignment Scores:

Pred. No.: 848 Length: 19237  
Score: 49.00 Matches: 10  
Percent Similarity: 80.00% Conservative: 2  
Best Local Similarity: 66.67% Mismatches: 3  
Query Match: 40.16% Indels: 0  
DB: 4 Gaps: 0

US-10-691-590-3 (1-26) x US-09-949-016-13666 (1-19237)

Qy 1 ArgArgGluGluGluGluSerGluGluGluThrPheGlyGlu 15

Db 17293 CGGAGGAGGAGGAAGGAAGCCGCCGAGGAGCGGGGAG 17249

## RESULT 12

US-09-489-039A-4753  
; Sequence 4753, Application US/09489039A  
; Patent No. 6610836

## GENERAL INFORMATION:

; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; PRIOR FILING DATE: 2000-01-27  
; PRIOR FILING DATE: 2000-01-27  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 4753  
; LENGTH: 330

; TYPE: DNA  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-4753

## Alignment Scores:

Pred. No.: 11.4 Length: 330  
Score: 48.00 Matches: 9  
Percent Similarity: 54.17% Conservative: 4  
Best Local Similarity: 37.50% Mismatches: 11  
Query Match: 39.34% Indels: 0  
DB: 4 Gaps: 0

US-10-691-590-3 (1-26) x US-09-489-039A-4753 (1-330)

Qy 3 GluGluGluGluGluSerGluGluGluThrPheGlyGluPhe\*\*\*GlnVal\*\*\*AlaPro 22

Db 220 GAACAGCATGCCAACAGGAGAATATGACTCTCGGGCAGCTCGCGCAGGTTCATCAGCGAC 279

Qy 23 Leu\*\*\*ProGly 26

Db 280 ATCTATCTCTGCG 291

## RESULT 13

US-09-489-039A-4835/c  
; Sequence 4835, Application US/09489039A  
; Patent No. 6610836

## GENERAL INFORMATION:

; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; PRIOR FILING DATE: 2000-01-27  
; PRIOR FILING DATE: 2000-01-27  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 4835  
; LENGTH: 420

; TYPE: DNA  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-4835

## Alignment Scores:

Pred. No.: 15 Length: 420  
Score: 48.00 Matches: 9  
Percent Similarity: 54.17% Conservative: 4  
Best Local Similarity: 37.50% Mismatches: 11  
Query Match: 39.34% Indels: 0  
DB: 4 Gaps: 0

US-10-691-590-3 (1-26) x US-09-489-039A-4835 (1-420)

Qy 3 GluGluGluGluGluSerGluGluGluThrPheGlyGluPhe\*\*\*GlnVal\*\*\*AlaPro 22

Db 192 GAACAGCATGCCAACAGGAGAATATGACTCTCGGGCAGCTCGCGCAGGTTCATCAGCGAC 133

Qy 23 Leu\*\*\*ProGly 26

Db 132 ATCTATCTCTGCG 121

## RESULT 14

US-08-998-416-421/c  
; Sequence 421, Application US/08998416  
; Patent No. 6239264

## GENERAL INFORMATION:

; APPLICANT: Philippsen, Peter  
; APPLICANT: Pohlmann, Rainer  
; APPLICANT: Steiner, Sabine  
; APPLICANT: Mohr, Christine  
; APPLICANT: Wendland, Jurgene  
; APPLICANT: Knechtle, Philipp  
; APPLICANT: Reibischung, Corinne  
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII  
; TITLE OF INVENTION: AND USES THEREOF  
; NUMBER OF SEQUENCES: 1152  
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 6239264artis Corporation

; STREET: 3054 Cornwallis Road

; CITY: Research Triangle Park

; STATE: No. 6239264th Carolina

; COUNTRY: USA

; ZIP: 27709

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/998,416

; FILING DATE: 24-DEC-1997

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: CH 0016/97

; FILING DATE: 31-DEC-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Weigs, J. Timothy

; REGISTRATION NUMBER: 38,241

; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 919-541-8587

; TELEFAX: 919-541-8689

; INFORMATION FOR SEQ ID NO: 421:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 790 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; ORIGINAL SOURCE:

; ORGANISM: PAG1311UP

US-08-998-416-421



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 5, 2005, 00:20:21 ; Search time 166.174 Seconds  
(without alignments)

926.218 Million cell updates/sec

Title: US-10-691-590-3

Perfect score: 122

Sequence: 1 RREQEESBEETGFEQXQVAPLXPG 26

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first: 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US10691590/runat\_02112005\_13134/app.query.fasta\_1.462  
-DB=N Geneseq 16Dec04 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCLALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10691590 @CGN 1\_1586 @runat\_02112005\_13134 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq 16Dec04:

- 1: Geneseqn1980s:
- 2: Geneseqn1990s:
- 3: Geneseqn2000s:
- 4: Geneseqn2001as:
- 5: Geneseqn2001bs:
- 6: Geneseqn2002as:
- 7: Geneseqn2002bs:
- 8: Geneseqn2003as:
- 9: Geneseqn2003bs:
- 10: Geneseqn2003cs:
- 11: Geneseqn2003ds:
- 12: Geneseqn2004as:
- 13: Geneseqn2004bs:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	117	95.9	1867	2	AQ20377 Sequence
C 2	65.5	53.7	6519	5	Abal6909 Human ner
C 3	59	48.4	1845	12	Ach91596 Human gen
4	59	48.4	2365	12	Adq18832 Human sof
5	59	48.4	2365	13	Adq38361 Human SNP

	59	48.4	2375	13	ADQ38360	Adq38360 Human SNP
	59	48.4	2382	12	ADQ23228	Adq23228 Human sof
C 8	59	48.4	4183	4	AAK74968	Aak74968 Human imm
C 9	59	48.4	4183	4	AAK75364	Aak75364 Human imm
10	57	46.7	10266	2	AAT33007	Aat33007 Mouse SRY
C 11	57	46.7	205388	12	ADQ97560	Adq97560 Mouse can
C 12	55	45.1	2589	4	ABL09646	Abi09646 Drosophil
13	54.5	44.7	4754	6	ABK52610	Abk52610 DNA encod
C 14	53	43.4	841	3	ABN81117	Abn81117 Shrimp po
15	53	43.4	4414	10	ADC30694	Adc30694 Human nov
16	53	43.4	4755	6	ABK52611	Abk52611 DNA encod
17	53	43.4	4804	5	ABV30199	Abv30199 Human pro
C 18	53	43.4	110000	12	ADN46845_11	Continuation (12 o
19	53	43.4	110000	12	ADN47591_08	Continuation (9 of
20	53	43.4	110000	12	ADN47591_09	Continuation (10 o
C 21	53	43.4	110000	12	ADN46123_11	Continuation (12 o
22	53	43.4	110000	12	ADN47209_08	Continuation (9 of
23	53	43.4	110000	12	ADN47209_09	Continuation (10 o
C 24	53	43.4	110000	12	ADN46464_11	Continuation (12 o
25	53	43.4	110000	12	ADN47960_08	Continuation (9 of
26	53	43.4	110000	12	ADN47960_09	Continuation (10 o
27	53	43.4	111206	11	ACN45152	Acn45152 Mouse gen
C 28	53	43.4	134841	11	ACN45172	Acn45172 Mouse gen
29	52	42.6	403	4	ABA08483	Abao8483 Human L-p
30	52	42.6	405	8	ABX62598	Abx62598 Arabidops
31	52	42.6	1845	12	ADO62521	Ado62521 Transcrip
32	52	42.6	3233	12	ADO62523	Ado62523 Transcrip
33	52	42.6	147309	6	ABK49450	Abk49450 Human tra
34	52	42.6	312477	12	ADP69744	Adp69744 Human ROC
35	51.5	42.2	6953	12	ADO39645	Ado39645 Human lam
C 36	51.5	42.2	13572	9	AAS58743	Aas58743 Human tra
37	51	41.8	939	5	AAS84137	Aas84137 DNA encod
C 38	51	41.8	5778	10	AAD55816	Aad55816 Micromono
C 39	51	41.8	17732	3	AAC68300	Aac68300 Lama2/APP
C 40	51	41.8	20623	3	AAC68294	Aac68294 Lama2/APP
C 41	51	41.8	22533	13	ADQ91713	Adq91713 Polyketid
C 42	51	41.8	31812	11	ACN44700	Acn44700 Mouse gen
43	51	41.8	58837	6	ABK52612	Abk52612 Human Cla
C 44	51	41.8	60196	10	AAD55810	Aad55810 Micromono
45	51	41.8	89736	13	ABD32966_10	Continuation (11 o

## ALIGNMENTS

RESULT 1  
AAQ20377  
ID AAQ20377 standard; cDNA; 1867 BP.  
XX  
AC AAQ20377;  
XX  
DT 16-APR-1992 (first entry)  
XX  
DE Sequence of 67 kD T. cacao protein cDNA.  
XX  
KW Cocoa; flavour; vicilin; seed storage protein; ss.  
XX  
OS Theobroma cacao.  
XX  
FH Key Location/Qualifiers  
FT CDS 14..1714  
FT /\*tag= a  
XX  
PN WO9119801-A.  
XX  
PD 26-DEC-1991.  
XX  
PF 11-JUN-1990; 90GB-00013016.  
XX  
PR 11-JUN-1990; 90GB-00013016.  
XX  
PA (MRSC ) MARS UK LTD.  
XX  
PI Spencer ME, Hodge R, Deakin EA, Ashton S;















KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis; ds.  
XX

OS Homo sapiens.

XX WO200157182-A2.

PN

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

25-SEP-2000; 2000US-0234997P.  
26-SEP-2000; 2000US-0234998P.  
26-SEP-2000; 2000US-0234999P.  
27-SEP-2000; 2000US-023484P.  
27-SEP-2000; 2000US-0235834P.  
27-SEP-2000; 2000US-0235836P.  
29-SEP-2000; 2000US-0236327P.  
29-SEP-2000; 2000US-0236367P.  
29-SEP-2000; 2000US-0236368P.  
29-SEP-2000; 2000US-0236369P.  
29-SEP-2000; 2000US-0236370P.  
29-SEP-2000; 2000US-0236802P.  
02-OCT-2000; 2000US-0237037P.  
02-OCT-2000; 2000US-0237038P.  
02-OCT-2000; 2000US-0237039P.  
13-OCT-2000; 2000US-0239935P.  
13-OCT-2000; 2000US-0239937P.  
20-OCT-2000; 2000US-0240960P.  
20-OCT-2000; 2000US-0241221P.  
20-OCT-2000; 2000US-0241785P.  
20-OCT-2000; 2000US-0241786P.  
20-OCT-2000; 2000US-0241787P.  
20-OCT-2000; 2000US-0241808P.  
20-OCT-2000; 2000US-0241809P.  
20-OCT-2000; 2000US-0241826P.  
01-NOV-2000; 2000US-0244617P.  
08-NOV-2000; 2000US-0246474P.  
08-NOV-2000; 2000US-0246475P.  
08-NOV-2000; 2000US-0246476P.  
08-NOV-2000; 2000US-0246477P.  
08-NOV-2000; 2000US-0246478P.  
08-NOV-2000; 2000US-0246523P.  
08-NOV-2000; 2000US-0246524P.  
08-NOV-2000; 2000US-0246525P.  
08-NOV-2000; 2000US-0246526P.  
08-NOV-2000; 2000US-0246527P.  
08-NOV-2000; 2000US-0246528P.  
08-NOV-2000; 2000US-0246532P.  
08-NOV-2000; 2000US-0246609P.  
08-NOV-2000; 2000US-0246610P.  
08-NOV-2000; 2000US-0246611P.  
08-NOV-2000; 2000US-0246613P.  
17-NOV-2000; 2000US-0249207P.  
17-NOV-2000; 2000US-0249208P.  
17-NOV-2000; 2000US-0249209P.  
17-NOV-2000; 2000US-0249210P.  
17-NOV-2000; 2000US-0249211P.  
17-NOV-2000; 2000US-0249212P.  
17-NOV-2000; 2000US-0249213P.  
17-NOV-2000; 2000US-0249214P.  
17-NOV-2000; 2000US-0249215P.  
17-NOV-2000; 2000US-0249216P.  
17-NOV-2000; 2000US-0249217P.  
17-NOV-2000; 2000US-0249218P.  
17-NOV-2000; 2000US-0249244P.  
17-NOV-2000; 2000US-0249245P.  
17-NOV-2000; 2000US-0249264P.  
17-NOV-2000; 2000US-0249265P.  
17-NOV-2000; 2000US-0249297P.  
17-NOV-2000; 2000US-0249299P.  
17-NOV-2000; 2000US-0249300P.  
01-DEC-2000; 2000US-0250160P.  
01-DEC-2000; 2000US-0250391P.  
05-DEC-2000; 2000US-0251030P.  
05-DEC-2000; 2000US-0251988P.  
05-DEC-2000; 2000US-0256719P.  
06-DEC-2000; 2000US-0251479P.  
08-DEC-2000; 2000US-0251856P.  
08-DEC-2000; 2000US-0251868P.  
08-DEC-2000; 2000US-0251869P.  
08-DEC-2000; 2000US-0251989P.  
08-DEC-2000; 2000US-0251990P.  
11-DEC-2000; 2000US-0254097P.

17-JAN-2001; 2001WO-US001354.

31-JAN-2000; 2000US-0179065P.

04-FEB-2000; 2000US-0180628P.

24-FEB-2000; 2000US-0184564P.

02-MAR-2000; 2000US-0186350P.

16-MAR-2000; 2000US-0189874P.

17-MAR-2000; 2000US-0190076P.

18-APR-2000; 2000US-0198123P.

19-MAY-2000; 2000US-0205515P.

07-JUN-2000; 2000US-0209467P.

28-JUN-2000; 2000US-0214886P.

30-JUN-2000; 2000US-0215135P.

07-JUL-2000; 2000US-0216647P.

07-JUL-2000; 2000US-0216880P.

11-JUL-2000; 2000US-0217487P.

11-JUL-2000; 2000US-0217496P.

14-JUL-2000; 2000US-0218290P.

26-JUL-2000; 2000US-0220963P.

26-JUL-2000; 2000US-0220964P.

14-AUG-2000; 2000US-0224518P.

14-AUG-2000; 2000US-0224519P.

14-AUG-2000; 2000US-022513P.

14-AUG-2000; 2000US-0225114P.

14-AUG-2000; 2000US-0225266P.

14-AUG-2000; 2000US-0225267P.

14-AUG-2000; 2000US-0225268P.

14-AUG-2000; 2000US-0225270P.

14-AUG-2000; 2000US-0225447P.

14-AUG-2000; 2000US-0225757P.

14-AUG-2000; 2000US-0225758P.

14-AUG-2000; 2000US-0225759P.

18-AUG-2000; 2000US-022679P.

22-AUG-2000; 2000US-0226681P.

22-AUG-2000; 2000US-0226868P.

22-AUG-2000; 2000US-0227182P.

23-AUG-2000; 2000US-0227009P.

30-AUG-2000; 2000US-0228924P.

01-SEP-2000; 2000US-0229287P.

01-SEP-2000; 2000US-0229343P.

01-SEP-2000; 2000US-0229344P.

05-SEP-2000; 2000US-0229345P.

05-SEP-2000; 2000US-0229509P.

06-SEP-2000; 2000US-0229513P.

06-SEP-2000; 2000US-0230437P.

08-SEP-2000; 2000US-0230438P.

08-SEP-2000; 2000US-0231242P.

08-SEP-2000; 2000US-0231243P.

08-SEP-2000; 2000US-0231244P.

08-SEP-2000; 2000US-0231413P.

08-SEP-2000; 2000US-0231414P.

08-SEP-2000; 2000US-0232080P.

12-SEP-2000; 2000US-0232081P.

14-SEP-2000; 2000US-02321968P.

14-SEP-2000; 2000US-0232397P.

14-SEP-2000; 2000US-0232398P.

14-SEP-2000; 2000US-0232399P.

14-SEP-2000; 2000US-0232400P.

14-SEP-2000; 2000US-0232401P.

14-SEP-2000; 2000US-0233063P.

14-SEP-2000; 2000US-0233064P.

14-SEP-2000; 2000US-0233065P.

21-SEP-2000; 2000US-0234223P.

21-SEP-2000; 2000US-0234274P.









CC (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628  
CC -ADC33394). The nucleic acids and polypeptides of the invention are  
CC useful in diagnostics, drug screening, forensics, gene mapping, in the  
CC identification of mutations responsible for genetic disorders or other  
CC traits, for assessing biodiversity, and in producing many other types of  
CC data and products dependent on DNA and amino acid sequences. They are  
CC also used for treating diseases such as Parkinson's disease, Alzheimer's  
CC disease and other neurodegenerative diseases, anaemia, platelet  
CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
CC cancer. The nucleic acids may also be used as hybridisation probes or  
CC primers, and in the recombinant production of a protein. The polypeptides  
CC are also useful in generating antibodies, as molecular weight markers,  
CC and as food supplements. The present sequence represents a specifically  
CC claimed human cDNA sequence of the invention. Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 4414 BP; 1497 A; 828 C; 1167 G; 922 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 134 Length: 4414  
Score: 53.00 Matches: 10  
Percent Similarity: 76.47% Conservative: 3  
Best Local Similarity: 58.82% Mismatches: 4  
Query Match: 43.44% Indels: 0  
DB: 10 Gaps: 0

US-10-691-590-3 (1-26) x ADC30694 (1-4414)

Qy 3 GluGlnGluGluSerGluGluGluThrPheGlyGluPhe\*\*\*GlnVal 19  
Db 3236 GACAGGAAGAGGAGACGAGGAGGAGGAATTGGAGACTTTCGGCTTGT 3286

Search completed: November 5, 2005, 03:06:51  
Job time : 193.174 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 5, 2005, 00:24:30 ; Search time 1280.78 Seconds  
(without alignments)  
983.645 Million cell updates/sec

Title: US-10-691-590-3  
Perfect score: 122  
Sequence: 1 RREQESEBETGTFQXVXAPLXPG 26

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgm2\_1/USP01.spool/US10691590/runat\_02112005\_120021\_13145/app\_query.fasta\_1.462  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -FRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10691590 @CGM\_1\_13851 @runat\_02112005\_120021\_13145 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=10 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_hgt.\*
- 3: gb\_in.\*
- 4: gb\_on.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	117	95.9	1867	6 A20606	A20606 67kD protei
2	117	95.9	3318	8 TCOSV	X62625 T.cacao csv
3	82	67.2	1380	8 TCOSVSV	X62626 T.cacao csv
4	66	54.1	242889	2 AC095192	AC095192 Rattus no

C	5	66	54.1	252743	2	AC136578	AC136578 Rattus no
C	6	65.5	53.7	37492	9	AC010649	AC010649 Homo sapi
C	7	61	50.0	171432	2	AC102322	AC102322 Mus muscu
C	8	60	49.2	214464	2	AC121397	AC121397 Rattus no
C	9	60	49.2	236863	2	AC097079	AC097079 Rattus no
C	10	60	49.2	242913	9	AC115994	AC115994 Homo sapi
C	11	59	48.4	2332	9	BC069795	BC069795 Homo sapi
C	12	59	48.4	2332	9	BC069802	BC069802 Homo sapi
C	13	59	48.4	2365	6	CQ715312	CQ715312 Sequence
C	14	59	48.4	2365	9	HUMHCPB	ME0052 Human histi
C	15	59	48.4	42003	2	AC145716	AC145716 Homo sapi
C	16	59	48.4	42920	2	AC145702	AC145702 Homo sapi
C	17	59	48.4	117628	9	AC008891	AC008891 Homo sapi
C	18	59	48.4	164635	2	AC143346	AC143346 Homo sapi
C	19	59	48.4	184579	2	AC105976	AC105976 Mus muscu
C	20	59	48.4	192387	10	AC151578	AC151578 Mus muscu
C	21	59	48.4	211763	10	AC123558	AC123558 Mus muscu
C	22	59	48.4	219784	10	AC112792	AC112792 Mus muscu
C	23	58	47.5	1828	9	AK129876	AK129876 Homo sapi
C	24	57	46.7	2490	8	MCPPC1A	X63774 M.cryptalli
C	25	57	46.7	10266	6	E11536	E11536 DNA from Y
C	26	57	46.7	163378	8	AP004087	AP004087 Oryza sat
C	27	57	46.7	191261	2	AC074304	AC074304 Mus muscu
C	28	57	46.7	224138	10	AC109243	AC109243 Mus muscu
C	29	57	46.7	229813	2	AC117613	AC117613 Mus muscu
C	30	56.5	46.3	221983	2	AC122881	AC122881 Mus muscu
C	31	56	45.9	64318	2	AC100364	AC100364 Mus muscu
C	32	56	45.9	144555	2	AC037451	AC037451 Homo sapi
C	33	56	45.9	166601	2	AC132753	AC132753 Rattus no
C	34	56	45.9	175629	9	AC097173	AC097173 Homo sapi
C	35	56	45.9	178209	2	AC150629	AC150629 Papio anu
C	36	56	45.9	178601	9	HSJ1022P6	AL109935 Human DNA
C	37	56	45.9	183935	2	AC112962	AC112962 Mus muscu
C	38	56	45.9	192562	10	AC132373	AC132373 Mus muscu
C	39	56	45.9	199875	9	AC134407	AC134407 Homo sapi
C	40	56	45.9	206630	2	AC150604	AC150604 Callithri
C	41	56	45.9	207404	2	AC149189	AC149189 Papio anu
C	42	56	45.9	216322	2	AC151686	AC151686 Gallus ga
C	43	56	45.9	235706	2	AC132182	AC132182 Rattus no
C	44	56	45.9	247402	10	AC125187	AC125187 Mus muscu
C	45	55.5	45.5	110000	1	RMES91985_02	Continuation (3 of

ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
A20606	67kD protein.	A20606	A20606.1	GI:579221	Theobroma cacao (cacao)	Theobroma cacao	Theobroma cacao
REFERENCE	1	(bases 1 to 1867)					
AUTHORS							
TITLE							
JOURNAL							
FEATURES							
source							
gene							
CDS							

RECOMBINANT 47 AND 31kD COCOA PROTEINS AND PRECURSOR

Patent: WO 9119801-A 6 26-DEC-1991;

Location/Qualifiers

1. .1867

/organism="Theobroma cacao"

/mol\_type="unassigned DNA"

/db\_xref="taxon:3641"

14. .1714

/gene="67kD protein"

14. .1714

/genes="67kD protein"

/codon\_start=1

/protein\_id="CAA01512.1"

/db\_xref="GI:579222"  
/translation="MVISKSPFIVLIFSLLSFALLSCGSVAYGRKQYERDPRQYEQ  
QOCRCSEATEREOEQCEQRCEREYKEQQOQBEELQRYOCCQRCQEQOQOQREQ  
QOCQRCWEQYKEQGEHYNHNKKNRSEEEQQRNNPYFPKRKSFQTRFRDEE  
GNFKILQRFAPNSPLKINDYRLAMFEANPNTFILPHCDAAEAYFVTNGKGTITFV  
THNKESYVQVGTVVSVAGSTVVVSDNOEKLTI AVLALPVNSPGKYLEFFPAGN  
NKPSYGCASFSEVLETVNTQREKLEILEORQKQOQOQGMFRKAKPQIRAIIS  
QOATSPRHGRGERLAINLLSQSPVYSNQNCRFFACPEDFSQFQNMVAVSAFKLNQ  
AIFVPHYNSKATFVFTDGYGQAQMACPHLSRQSQSGRQDRREBESEETFG  
EFQOVKAPLSFGDVFVAPAGHAVTFFASKDQPLNAVAFGLNAQNNQRIPLAGRPPLN  
OMDSEAKELSGVPSKLVNDIIFNNPDES YFMSFQQRQDRRDRGNPLASILDFARLF  
"

ORIGIN

Alignment Scores:  
Pred. No.: 3,68e-11 Length: 1867  
Score: 117.00 Matches: 23  
Percent Similarity: 88.46% Conservative: 0  
Best Local Similarity: 88.46% Mismatches: 3  
Query Match: 95.90% Indels: 0  
DB: 6 Gaps: 0

US-10-691-590-3 (1-26) x A20606 (1-1867)

QY 1 ArgArgGluGluGluSerGluGluThrPheGlyGluPhe\*\*\*GlnVal\*\*\* 20  
Db 1322 AGAAGAGACAAGAGAGAGTTCAGAGAGGAGACATTGGAGAAATTCACGAGGTCAAA 1381  
QY 21 AlaProLeu\*\*\*ProGly 26  
Db 1382 GCCCATTCACCTGGT 1399

RESULT 2

TCCSV 3318 bp DNA linear PLN 03-DEC-1993  
LOCUS T.cacao csv gene for seed vicilin.  
DEFINITION X62625.38078  
ACCESSION X62625.1 GI:21910  
VERSION csv gene; seed protein; vicilin.  
KEYWORDS Theobroma cacao (cacao)  
SOURCE Theobroma cacao  
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;  
Theobroma.

REFERENCE 1 (bases 1 to 3318)  
AUTHORS McHenry, L. and Fritz, P.J.  
TITLE Comparison of the structure and nucleotide sequences of vicilin  
genes of cocoa and cotton raise questions about vicilin evolution  
JOURNAL Plant Mol. Biol. 18 (6), 1173-1176 (1992)  
MEDLINE 92288309  
PUBMED 1600151  
REFERENCE 2 (bases 1 to 3318)  
AUTHORS McHenry, L.  
TITLE Direct Submission  
JOURNAL Submitted (15-OCT-1991) L. McHenry, Pennsylvania State University,  
111 Borland Lab, University Park, PA 16802, USA  
COMMENT See also x62626  
Overlap of sequenced fragments.

FEATURES

source  
1..3318  
/organism="theobroma cacao"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:3641"  
/clone="PHDSP1.7, PHDSP1.2, PHDSE4.5"  
/tissue\_type="leaves"  
/clone\_lib="cocoa genomic library in lambda GEM II"  
/dev\_stage="mature tree"  
TATA\_signal 543..549  
gene 610..2581  
/notes="csv gene"  
/genes="csv"  
join(610..1240,1331..1506,1597..1677,1780..2064,  
CDS

2177..2581)  
/gene="csv"  
/codon\_start=1  
/product="vicilin"  
/protein\_id="CAA44493.1"  
/db\_xref="GI:21911"  
/db\_xref="GOA:Q43358"  
/db\_xref="UniProt/Swiss-Prot:Q43358"  
/translation="MVISKSPFIVLIFSLLSFALLSCGSVAYGRKQYERDPRQYEQ  
QOCRCSEATEREOEQCEQRCEREYKEQQOQBEELQRYOCCQRCQEQOQOQREQ  
QOCQRCWEQYKEQGEHYNHNKKNRSEEEQQRNNPYFPKRKSFQTRFRDEE  
GNFKILQRFAPNSPLKINDYRLAMFEANPNTFILPHCDAAEAYFVTNGKGTITFV  
THNKESYVQVGTVVSVAGSTVVVSDNOEKLTI AVLALPVNSPGKYLEFFPAGN  
NKPSYGCASFSEVLETVNTQREKLEILEORQKQOQOQGMFRKAKPQIRAIIS  
QOATSPRHGRGERLAINLLSQSPVYSNQNCRFFACPEDFSQFQNMVAVSAFKLNQ  
AIFVPHYNSKATFVFTDGYGQAQMACPHLSRQSQSGRQDRREBESEETFG  
EFQOVKAPLSFGDVFVAPAGHAVTFFASKDQPLNAVAFGLNAQNNQRIPLAGRPPLN  
HKQNTNVIKFTVKASAY"  
/gene="csv"  
join(682..1240,1331..1506,1597..1677,1780..2064,  
2177..2578)  
/gene="csv"  
/product="vicilin"  
/gene="csv"  
/number=1  
/gene="csv"  
1241..1330  
/gene="csv"  
/number=1  
1331..1506  
/gene="csv"  
/number=2  
1507..1596  
/gene="csv"  
/number=2  
1597..1677  
/gene="csv"  
/number=3  
1678..1779  
/gene="csv"  
/number=3  
1780..2064  
/gene="csv"  
/number=4  
2065..2176  
/gene="csv"  
/number=4  
2177..2582  
/number=5  
2856..2861

sig\_peptide

mat\_peptide

exon

intron

exon

intron

exon

intron

exon

intron

exon

polyA\_signal  
ORIGIN

Alignment Scores:

Pred. No.: 6,82e-11 Length: 3318  
Score: 117.00 Matches: 23  
Percent Similarity: 88.46% Conservative: 0  
Best Local Similarity: 88.46% Mismatches: 3  
Query Match: 95.90% Indels: 0  
DB: 8 Gaps: 0

US-10-691-590-3 (1-26) x TCCSV (1-3318)

QY 1 ArgArgGluGluGluSerGluGluThrPheGlyGluPhe\*\*\*GlnVal\*\*\* 20  
Db 2312 AGAAGAGACAAGAGAGTTCAGAGAGGAGACATTGGAGAAATTCACGAGGTCAAA 2371

QY 21 AlaProLeu\*\*\*ProGly 26

Db 2372 GCCCATTCACCTGGT 2389

RESULT 3

TCCSVS

LOCUS	TCCSVSV	1380 bp	mRNA	linear	PLN 06-DEC-1992
DEFINITION	T.cacaoa csv mRNA for seed vicilin.				
ACCESSION	X62626				
VERSION	X62626.1 GI:21912				
KEYWORDS	csv gene; seed protein; vicilin.				
SOURCE	Theobroma cacao (cacao)				
ORGANISM	Theobroma cacao				
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae; Theobroma.					
REFERENCE	1 (bases 1 to 1380) McHenry,L. and Fritz,P.J. Comparison of the structure and nucleotide sequences of vicilin genes of cocoa and cotton raise questions about vicilin evolution Plant Mol. Biol. 18 (6), 1173-1176 (1992)				
JOURNAL	MEDLINE				
PUBMED	92288309				
REFERENCE	2 (bases 1 to 1380) McHenry,L. Direct Submission Submitted (15-OCT-1991) L. McHenry, Pennsylvania State University, 111 Borland Lab, University Park, PA 16802, USA				
JOURNAL	TITILE				
COMMENT	See also x62625 Overlap of sequenced fragments, sequence partial. Location/Qualifiers 1..1380 /organism="Theobroma cacao" /mol_type="mRNA" /db_xref="taxon:3641" /clone="pHT4C3" /tissue_type="cotyledon" /clone_lib="cocoa seed library in lambda gt10" /dev_stage="130DAP" 1..1380 /gene="csv" <1..>1380 /gene="csv" /product="vicilin" /protein_id="CAA4494.1" /db_xref="GI:21913" /db_xref="GOA:Q43358" /db_xref="UniProt/Swiss-Prot:Q43358" /translation="POIKSIANMVISKSPFIVLIFSLLLGALLCSGVSAVGRKOVER DRQYEQCORECESEATEEEOCEOCEREYKEOQOEEELQRYOQCOGRQOE QQOQREOQOQCKWEQYKQGERGEHYNHKNRSEEGQQRNNPYFKRSP QTRFDEGNFKILQFAENSPPLKINDYRLAMPNPNTFILPHCDAAEALFVFN QRTFDITVENKESYNQRTVTVSPAGSTVYVSDNQEKUATLIALVFNPSGKY ELFPAGNKNPSYVYGASVEYLVFTVQREKLEILEEQRQOQOQGMFRKAK PQIRAIQOATSPRRHGERLAINLSQSPVSNQNRPFACPFDSFOFNMDVAV SAFLKNQGAIFVPHYNSKATFVFTDGYGVAQACPHLSRQSQSQSGRQDRREQEE ESEETFTGBF" <1..24 /gene="csv" 25..1380 /gene="csv" /product="vicilin"				
gene					
CDS					
sig_peptide					
mat_peptide					
ORIGIN					
Alignment Scores:					
Pred. No.:	6.08e-05	Length:	1380		
Score:	82.00	Matches:	16		
Percent Similarity:	100.00%	Conservative:	0		
Best Local Similarity:	100.00%	Mismatches:	0		
Query Match:	67.21%	Indels:	0		
DB:	8	Gaps:	0		
US-10-691-590-3 (1-26) x TCCSVSV (1-1380)					
Qy	1 ArgArgGluGlnGluGluSerGluGluGluThrPhrGlyGluPhe 16				
Qb	1333 AGAAGAGACAGACAGACAGATTCAGAGAGGAGACATTTCGAGATTC 1380				









```

* 31239 34733: contig of 3495 bp in length
* 34734 34833: gap of 100 bp
* 34834 40484: contig of 5651 bp in length
* 40485 40584: gap of 100 bp
* 40585 45923: contig of 5338 bp in length
* 45923 46023: gap of 100 bp
* 46023 51563: contig of 5540 bp in length
* 51563 51663: gap of 100 bp
* 51663 58135: contig of 6473 bp in length
* 58135 58235: gap of 100 bp
* 58235 68445: contig of 10210 bp in length
* 68445 68545: gap of 100 bp
* 68545 79090: contig of 10545 bp in length
* 79090 79191: gap of 100 bp
* 79191 92907: contig of 13717 bp in length
* 92907 93007: gap of 100 bp
* 93007 110376: contig of 17369 bp in length
* 110376 110476: gap of 100 bp
* 110476 110477: gap of 100 bp
* 110477 171432: contig of 60956 bp in length.

```

## FEATURES

```

source
1..171432
  /organism="Mus musculus"
  /mol_type="genomic DNA"
  /db_xref="taxon:10090"
  /clone="RP24-462C5"
  /clone_lib="RPCI-24 Male Mouse BAC"
1..681
  /note="assembly_fragment"
782..1476
  /note="assembly_fragment"
1577..2062
  /note="assembly_fragment"
2163..2595
  /note="assembly_fragment"
2696..3398
  /note="assembly_fragment"
3499..4884
  /note="assembly_fragment"
4985..5587
  /note="assembly_fragment"
5688..6845
  /note="assembly_fragment"
6946..7876
  /note="assembly_fragment"
7977..9201
  /note="assembly_fragment"
9302..11037
  /note="assembly_fragment"
11138..13003
  /note="assembly_fragment"
13104..15285
  /note="assembly_fragment"
15386..17319
  /note="assembly_fragment"
17420..19929
  /note="assembly_fragment"
20030..22691
  /note="assembly_fragment"
22792..25794
  /note="assembly_fragment"
25895..28135
  /note="assembly_fragment"
28236..31138
  /note="assembly_fragment"
31239..34733
  /note="assembly_fragment"
34834..40484
  /note="assembly_fragment"

```

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

```

Best Local Similarity: 52.17%
Query Match: 50.00%
DB: 2
Gaps: 0

```

US-10-691-590-3 (1-26) x AC102322 (1-171432)

```

Qy 3 GluGluGluGluSerGluGluGluThrPheGlyClyuPhe***GlnVal***AlaPro 22
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 112717 GAAGAAGAGAGAGAAGAAGAAGAAGAATTTGGAGAGCTGGCAGCGTAGTCATCGCT 112776

```

Qy 23 Leu\*\*\*Pro 25

```

    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 112777 GTGATCCCA 112785

```

RESULT 8

AC121397

LOCUS

DEFINITION

AC121397.4 GI:25008414

HTG: HTGS PHASE1; HTGS DRAFT; HTGS\_ENRICHED.

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE

AUTHORS

Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,

Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,

Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,

Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,

Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,

Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,

Cardenas, V., Carter, K., Cavazos, J., Cesar, H., Center, A.,

Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,

Chleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,

Davila, M. L., Davis, C., Davy-Carrroll, L., De Anda, C., Dederich, D.,

Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,

Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,

Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,

Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,

Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,

Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,

Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,

Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,

Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hughes, M.,

Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,

Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,

Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,

Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,

Lorensueta, L., Louised, H., Lofado, R. J., Lu, X., Ma, J.,

Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,

Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,

Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,

Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,

Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,

Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,

Nwankwelu, O., Okwuonu, G., Olarinsunagoo, A., Pal, S., Parks, K.,

Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,

Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L. L.,

Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,

Rives, C., Rodkey, T., Rojars, A., Rose, M., Rose, R., Ruiz, S. J.,

Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,

Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajes, D.,

Sheed, A., Sodergren, E., Song, X. Z., Sorelle, R., Sosa, J.,

Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,

Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,

Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,

Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,

Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K.,

Worley, K.,

Worley, K.,

Worley, K.





```

TITLE
JOURNAL
REFERENCE
AUTHORS
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 242913)
Birren,B., Nussbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavsky,L., Bouckigalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collimore,A., Cook,A., Cooke,P., Corum,B., DeArrellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Menus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (31-JAN-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 242913)
Birren,B., Nussbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavsky,L., Bouckigalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collimore,A., Cook,A., Cooke,P., Corum,B., DeArrellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Menus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (01-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 1, 2003 this sequence version RepeatMasker
All repeats were identified using RepeatMasker:
Smith, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L26130
Center clone name: 2535_L_24
----- Location/Qualifiers
1. .242913
/organisms="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
FEATURES
source
/clone_lib="CITD2 Human BAC"
/clone="CTD-2535L24"
/chromosome="17"
/map="17"
/complement(1..80)
/rpt_family="Alu/FLAM"
/complement(250..553)
/rpt_family="Alu"
/complement(1209..1510)
/rpt_family="AluSq"
1595..1903
/rpt_family="AluSq"
/complement(2033..2195)
/rpt_family="MLT1P2"
/complement(2233..2425)
/rpt_family="MLT1P2"
/complement(2601..2650)
/rpt_family="L2"
2859..3173
/rpt_family="AluSq"
/complement(3546..3842)
/rpt_family="AluSq"
4125..4432
/rpt_family="AluSq"
4448..4471
/rpt_family="AT rich"
/complement(4765..5063)
/rpt_family="AluSq"
5482..5595
/rpt_family="MIR3"
/complement(5596..5905)
/rpt_family="AluSq"
5906..5961
/rpt_family="MIR3"
/complement(6188..6311)
/rpt_family="MIR"
/complement(7068..7373)
/rpt_family="AluSp"
7692..7737
/rpt_family="(TG)n"
/complement(8354..8637)
/rpt_family="AluSq"
/complement(8853..9158)
/rpt_family="AluSq"
10104..10140
/rpt_family="(CGGGG)n"
10153..10182
/rpt_family="GC rich"
10328..10358
/rpt_family="GC rich"
10435..10581
/rpt_family="(CGGGG)n"
10723..10750
/rpt_family="GC rich"
10882..10950
/rpt_family="GC rich"
11388..11463
/rpt_family="C-rich"
11566..11603
/rpt_family="GC rich"
11691..11732
/rpt_family="GC rich"
16081..16232
/rpt_family="MIR"
17557..17600
/rpt_family="(GGA)n"
/complement(18583..18760)
/rpt_family="L2"
/complement(18771..19079)
/rpt_family="AluSp"
/complement(19354..19677)
/rpt_family="L3b"
/complement(19763..19879)

```











GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 5, 2005, 02:54:27 ; Search time 900.835 Seconds  
(without alignments)  
817.035 Million cell updates/sec

Title: US-10-691-590-1  
Perfect score: 458  
Sequence: 1 RREQEESEETFGFQQVK.....FLNHKQNTNVIKFTVKASAY 89

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 9794790 seqs, 4134909567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USP01.spool/US10691590/runat\_02112005\_120023\_13233/app\_query.fasta\_1.462  
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DLOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US10691590@cgn\_1\_1\_582@runat\_02112005\_120023\_13233  
-NCPU=6 -ICPU=3 -NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=130 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA: \*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq: \*  
2: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq: \*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq: \*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq: \*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq: \*  
6: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq: \*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq: \*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq: \*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq: \*  
10: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq: \*  
11: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq: \*  
12: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq: \*  
13: /cgn2\_6/ptodata/2/pubpna/US09D\_PUBCOMB.seq: \*  
14: /cgn2\_6/ptodata/2/pubpna/US09E\_PUBCOMB.seq: \*  
15: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq: \*  
16: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq: \*  
17: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq: \*  
18: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq: \*  
19: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq: \*  
20: /cgn2\_6/ptodata/2/pubpna/US10E\_PUBCOMB.seq: \*  
21: /cgn2\_6/ptodata/2/pubpna/US10F\_PUBCOMB.seq: \*  
22: /cgn2\_6/ptodata/2/pubpna/US10G\_PUBCOMB.seq: \*  
23: /cgn2\_6/ptodata/2/pubpna/US10H\_PUBCOMB.seq: \*  
24: /cgn2\_6/ptodata/2/pubpna/US10I\_PUBCOMB.seq: \*  
25: /cgn2\_6/ptodata/2/pubpna/US10J\_PUBCOMB.seq: \*  
26: /cgn2\_6/ptodata/2/pubpna/US11\_PUBCOMB.seq: \*  
27: /cgn2\_6/ptodata/2/pubpna/US11A\_PUBCOMB.seq: \*  
28: /cgn2\_6/ptodata/2/pubpna/US11B\_PUBCOMB.seq: \*  
29: /cgn2\_6/ptodata/2/pubpna/US11C\_PUBCOMB.seq: \*  
30: /cgn2\_6/ptodata/2/pubpna/US11D\_PUBCOMB.seq: \*  
31: /cgn2\_6/ptodata/2/pubpna/US11E\_PUBCOMB.seq: \*  
32: /cgn2\_6/ptodata/2/pubpna/US11F\_PUBCOMB.seq: \*  
33: /cgn2\_6/ptodata/2/pubpna/US11G\_PUBCOMB.seq: \*  
34: /cgn2\_6/ptodata/2/pubpna/US11H\_PUBCOMB.seq: \*  
35: /cgn2\_6/ptodata/2/pubpna/US11I\_PUBCOMB.seq: \*  
36: /cgn2\_6/ptodata/2/pubpna/US11J\_PUBCOMB.seq: \*  
37: /cgn2\_6/ptodata/2/pubpna/US11K\_PUBCOMB.seq: \*  
38: /cgn2\_6/ptodata/2/pubpna/US11L\_PUBCOMB.seq: \*  
39: /cgn2\_6/ptodata/2/pubpna/US11M\_PUBCOMB.seq: \*  
40: /cgn2\_6/ptodata/2/pubpna/US11N\_PUBCOMB.seq: \*  
41: /cgn2\_6/ptodata/2/pubpna/US11O\_PUBCOMB.seq: \*  
42: /cgn2\_6/ptodata/2/pubpna/US11P\_PUBCOMB.seq: \*  
43: /cgn2\_6/ptodata/2/pubpna/US11Q\_PUBCOMB.seq: \*  
44: /cgn2\_6/ptodata/2/pubpna/US11R\_PUBCOMB.seq: \*  
45: /cgn2\_6/ptodata/2/pubpna/US11S\_PUBCOMB.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	159	34.7	562	20	US-10-021-323-762 Sequence 762, Appl
2	156.5	34.2	2032	10	US-09-731-375A-1 Sequence 1, Appli
3	156.5	34.2	2032	15	US-10-228-806-1 Sequence 1, Appli
4	156.5	34.2	2032	18	US-10-100-303A-5 Sequence 5, Appli
5	156.5	34.2	2032	21	US-10-728-323-1 Sequence 1, Appli
6	156.5	34.2	2032	21	US-10-728-051-1 Sequence 1, Appli
7	156.5	34.2	2032	22	US-10-898-551-1 Sequence 1, Appli
8	156.5	34.2	2032	22	US-10-958-324-5 Sequence 5, Appli
9	154.5	33.7	498	20	US-10-021-323-2828 Sequence 2828, Ap
C 10	151.5	33.1	2003	20	US-10-437-963-43667 Sequence 43667, A
11	151	33.0	1949	18	US-10-100-303A-6 Sequence 6, Appli
12	145.5	31.8	1587	19	US-10-425-114-7095 Sequence 7095, Ap
13	145.5	31.8	1933	19	US-10-425-114-23659 Sequence 23659, A
14	145.5	31.8	2001	19	US-10-425-114-3415 Sequence 3415, Ap
15	145.5	31.8	2926	21	US-10-425-115-117618 Sequence 117618, A
16	143	31.2	614	19	US-10-425-114-8131 Sequence 8131, Ap
17	143	31.2	720	19	US-10-425-114-15167 Sequence 15167, A
18	143	31.2	1037	19	US-10-425-114-14858 Sequence 14858, A
19	143	31.2	1107	19	US-10-425-114-8117 Sequence 8117, Ap
20	143	31.2	1856	19	US-10-425-114-9571 Sequence 9571, Ap
C 21	143	31.2	5642	19	US-10-424-598-10364 Sequence 10364, A
22	142	31.0	1920	9	US-09-758-652-2 Sequence 2, Appli
23	142	31.0	1920	19	US-10-684-651-2 Sequence 2, Appli
24	142	31.0	1920	20	US-10-757-074-2 Sequence 2, Appli
25	142	31.0	1920	20	US-10-757-155-2 Sequence 2, Appli
26	142	31.0	1920	20	US-10-757-667-2 Sequence 2, Appli
27	140.5	30.7	2171	9	US-09-331-631A-4 Sequence 4, Appli
28	140.5	30.7	2171	17	US-10-147-095-4 Sequence 4, Appli
29	140	30.6	1254	18	US-10-245-227B-15 Sequence 15, Appl
30	140	30.6	1278	18	US-10-245-227B-13 Sequence 13, Appl
31	140	30.6	1320	9	US-09-758-652-3 Sequence 3, Appli
32	140	30.6	1320	18	US-10-245-227B-3 Sequence 3, Appli
33	140	30.6	1320	19	US-10-684-651-3 Sequence 3, Appli
34	140	30.6	1320	20	US-10-757-074-3 Sequence 3, Appli
35	140	30.6	1320	20	US-10-757-155-3 Sequence 3, Appli
36	140	30.6	1320	20	US-10-757-667-3 Sequence 3, Appli
37	137	29.9	342	19	US-10-425-114-8154 Sequence 8154, Ap
38	137	29.9	536	19	US-10-425-114-7591 Sequence 7591, Ap
39	137	29.9	641	19	US-10-425-114-8571 Sequence 8571, Ap
40	137	29.9	673	19	US-10-425-114-7653 Sequence 7653, Ap
41	137	29.9	863	19	US-10-425-114-7590 Sequence 7590, Ap
42	137	29.9	880	19	US-10-425-114-7667 Sequence 7667, Ap
43	137	29.9	934	19	US-10-425-114-6929 Sequence 6929, Ap
44	137	29.9	951	19	US-10-425-114-8579 Sequence 8579, Ap
45	137	29.9	1011	19	US-10-425-114-8555 Sequence 8555, Ap

# ALIGNMENTS

## RESULT 1

US-10-021-323-762/c  
; Sequence 762, Application US/10021323  
; Publication No. US20040123340A1  
; GENERAL INFORMATION:  
; APPLICANT: Deikman, Jill  
; APPLICANT: Feng, Paul C.C.  
; APPLICANT: Fincher, Karen L.  
; APPLICANT: Ziegler, Todd E.  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(52274)B  
; CURRENT APPLICATION NUMBER: US/10/021.323  
; CURRENT FILING DATE: 2001-12-12

[illegible]









Db 1975 GCCGACAGGAACGAGAAGGTGTTCTGCGCGC 2007

Search completed: November 5, 2005, 06:20:50  
Job time : 907.835 secs

**This Page Blank (uspto)**